



**ÖZ-DİKKAT MEKANİZMASI REHBERLİĞİNDE
İSKELET TABANLI EYLEM TANIMA İÇİN
UZAYSAL-ZAMANSAL ÇİZGE EVRİŞİMLİ AĞ
MİMARİLERİNİN GELİŞTİRİLMESİ**

Uğur KILIÇ

Danışman: Dr. Öğr. Üyesi Gülşah TÜMÜKLÜ ÖZYER

Doktora Tezi

Bilgisayar Mühendisliği Ana Bilim Dalı

2025

(Her hakkı saklıdır.)

T.C.
ATATÜRK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANA BİLİM DALI

**ÖZ-DİKKAT MEKANİZMASI REHBERLİĞİNDE İSKELET TABANLI EYLEM
TANIMA İÇİN UZAYSAL-ZAMANSAL ÇİZGE EVRİŞİMLİ AĞ MİMARİLERİNİN
GELİŞTİRİLMESİ**

(Development of Spatio-Temporal Graph Convolutional Network Architectures for Skeleton-
Based Action Recognition Guided by Self-Attention Mechanism)

DOKTORA TEZİ

Uğur KILIÇ

Danışman: Dr. Öğr. Üyesi Gülşah TÜMÜKLÜ ÖZYER
İkinci Tez Danışmanı: Dr. Öğr. Üyesi Özge ÖZTİMUR KARADAĞ

Erzurum
Haziran, 2025

T.C.
ATATÜRK ÜNİVERSİTESİ
Fen Bilimleri Enstitüsü Müdürlüğü

TEZ KABUL VE ONAY TUTANAĞI

**ÖZ-DİKKAT MEKANİZMASI REHBERLİĞİNDE İSKELET TABANLI EYLEM
TANIMA İÇİN UZAYSAL-ZAMANSAL ÇİZGE EVRİŞİMLİ AĞ MİMARİLERİNİN
GELİŞTİRİLMESİ**

Dr. Öğr. Üyesi Gülşah TÜMÜKLÜ ÖZYER danışmanlığında, Uğur KILIÇ tarafından hazırlanan bu çalışma, 27/05/2024 tarihinde aşağıdaki jüri tarafından Bilgisayar Mühendisliği Anabilim Dalı Bilgisayar Mühendisliği Bilim Dalı'nda Doktora tezi olarak **oybirliği** ile kabul edilmiştir.

Jüri Başkanı:	Prof. Dr. Şeref Naci ENGİN <i>Yıldız Teknik Üniversitesi</i>	Aslı Islak İmzalıdır
Danışman:	Dr. Öğr. Üyesi Gülşah TÜMÜKLÜ ÖZYER <i>Atatürk Üniversitesi</i>	Aslı Islak İmzalıdır
Jüri Üyesi:	Prof. Dr. Sinan KALKAN <i>Orta Doğu Teknik Üniversitesi</i>	Aslı Islak İmzalıdır
Jüri Üyesi:	Prof. Dr. Muhammed Emre KESKİN <i>Atatürk Üniversitesi</i>	Aslı Islak İmzalıdır
Jüri Üyesi:	Doç. Dr. Ferhat BOZKURT <i>Atatürk Üniversitesi</i>	Aslı Islak İmzalıdır
Jüri Üyesi:	Dr. Öğr. Üyesi Ahmet COŞKUNÇAY <i>Atatürk Üniversitesi</i>	Aslı Islak İmzalıdır
İkinci Tez Danışmanı	Dr. Öğr. Üyesi Özge ÖZTİMUR KARADAĞ <i>Alanya Alaaddin Keykubat Üniversitesi</i>	Aslı Islak İmzalıdır

Bu tezin Atatürk Üniversitesi Lisansüstü Eğitim ve Öğretim Yönetmeliği'nin ilgili maddelerinde belirtilen şartları yerine getirdiğini onaylarım.

Prof.Dr. Alper NUHOĞLU

Enstitü Müdürü

Aslı Islak İmzalıdır

Bu çalışma TÜBİTAK 1002-A kapsamında 124E309 numaralı proje ile, TÜBİTAK 3501 kapsamında 123E635 numaralı proje ile ve FDK-2023-11957 kodlu BAP projesi kapsamında Atatürk Üniversitesi tarafından desteklenmiştir.

Not: Bu tezde kullanılan özgün ve başka kaynaklardan yapılan bildiriş, çizelge, şekil ve fotoğrafların kaynak olarak kullanımı, 5846 sayılı Fikir ve Sanat Eserleri Kanunundaki hükümlere tabidir.

ETİK BİLDİRİM VE İNTİHAL BEYAN FORMU

Doktora Tezi olarak Dr. Öğr. Üyesi Gülşah TÜMÜKLÜ ÖZYER danışmanlığında sunulan “Öz-Dikkat Mekanizması Rehberliğinde İskelet Tabanlı Eylem Tanıma İçin Uzaysal-Zamansal Çizge Evrişimli Ağ Mimarilerinin Geliştirilmesi” başlıklı çalışmanın tarafımızdan bilimsel etik ilkelere uyularak yazıldığını, yararlanılan eserlerin kaynakçada gösterildiğini, Fen Bilimleri Enstitüsü tarafından belirlenmiş olan Turnitin Programı benzerlik oranlarının aşılmadığını ve aşağıdaki oranlarda olduğunu beyan ederiz.

Tez Bölümleri	Tezin Benzerlik Oranı (%)	Maksimum Oran (%)
Giriş	12	30
Kuramsal Temeller	20	30
Materyal ve Metot	22	35
Araştırma Bulguları ve Tartışma	11	20
Sonuçlar	2	20
Tezin Geneli	17	25

Not: Yedi kelimeye kadar benzerlikler ile Başlık, Kaynakça, İçindekiler, Teşekkür, Dizin ve Ekler kısımları tarama dışı bırakılabilir. Yukarıdaki azami benzerlik oranları yanında tek bir kaynaktan olan benzerlik oranlarının %5'den büyük olmaması gerekir.

Beyan edilen bilgilerin doğru olduğunu, aksi halde doğacak hukuki sorumlulukları kabul ve beyan ederiz.

Tez Yazarı (Öğrenci)	Tez Danışmanı
Uğur KILIÇ	Dr. Öğr. Üyesi Gülşah TÜMÜKLÜ ÖZYER
13.6.2025	13.6.2025
İmza: Aslı Islak İmzalıdır	İmza: Aslı Islak İmzalıdır

* Tez ile ilgili YÖKTEZ’de yayınlamasına ilişkin bir engelleme var ise aşağıdaki alanı doldurunuz.

☐ Tezle ilgili patent başvurusu yapılması / patent alma sürecinin devam etmesi sebebiyle Enstitü Yönetim Kurulunun/....../.... tarih ve sayılı kararı ile teze erişim 2 (iki) yıl süreyle engellenmiştir.

☐ Enstitü Yönetim Kurulunun/....../.... tarih ve sayılı kararı ile teze erişim 6 (altı) ay süreyle engellenmiştir.

TEŞEKKÜR

Bu tezin konusu ve içeriğinin şekillenmesinde değerli bilgi ve deneyimleriyle yol gösteren, her zaman destek ve teşviklerini esirgemeyen danışman hocam Sayın Dr. Öğr. Üyesi Gülşah TÜMÜKLÜ ÖZYER'e en içten teşekkürlerimi sunarım. Çalışmam süresince sağladığı kıymetli katkılar ve yönlendirmeler için ikinci danışmanım Sayın Dr. Öğr. Üyesi Özge ÖZTİMUR KARADAĞ'a da gönülden teşekkür ederim.

Ayrıca tez izleme jürimde yer alan, bilgi ve deneyimleriyle çalışmama önemli katkılarda bulunan Sayın Dr. Öğr. Üyesi Ahmet COŞKUNÇAY'a ve Sayın Prof. Dr. Muhammed Emre KESKİN'e değerli görüşleri ve destekleri için teşekkür ederim.

Her zaman yanımda olan ve desteğini hiçbir zaman esirgemeyen aileme; rahmetli babama, rahmetli anneme ve kardeşlerime sonsuz teşekkür ederim.

Bu çalışmaya verdikleri maddi desteklerden dolayı, 124E309 ve 123E635 numaralı projeler kapsamında Türkiye Bilimsel ve Teknolojik Araştırma Kurumu'na (TÜBİTAK) ve FDK-2023-11957 kodlu BAP projesi kapsamında Atatürk Üniversitesi'ne sağladıkları destekler için teşekkür ederim.

Uğur KILIÇ

ÖZET

DOKTORA TEZİ

ÖZ-DİKKAT MEKANİZMASI REHBERLİĞİNDE İSKELET TABANLI EYLEM TANIMA İÇİN UZAYSAL-ZAMANSAL ÇİZGE EVRİŞİMLİ AĞ MİMARİLERİNİN GELİŞTİRİLMESİ

Uğur KILIÇ

Danışman: Dr. Öğr. Üyesi Gülşah TÜMÜKLÜ ÖZYER
İkinci Tez Danışmanı: Dr. Öğr. Üyesi Özge ÖZTİMUR KARADAĞ

Amaç: İskelet verileri, çevresel değişikliklere karşı dayanıklılığı, hesaplama verimliliği, kompakt yapısı ve mahremiyet endişelerini minimize etmeleri gibi avantajları nedeniyle son yıllarda eylem tanıma önemli bir modalite haline gelmiştir. Derin öğrenmenin gelişmesiyle birlikte, iskelet verilerini kullanarak eylem tanıma yapan birçok yöntem geliştirilmiştir. Özellikle, iskelet verilerinin topolojik yapısının çizge olarak modellenmeye uygun olması sebebiyle uzaysal-zamansal çizge evrışımsel ağlara (ST-GCN) olan ilgi artmıştır. Fakat ST-GCN modelleri sabit çizge topolojisi sınırlamaları, uzak mesafeli eklemeler arasındaki bağımlılıkların yetersiz temsili, uzun menzilli zamansal bağımlılıkların eksikliği ve koordineli hareketlerin yetersiz temsili gibi yapısal sınırlamalar devam etmektedir. Bu doktora çalışmasında, ST-GCN modellerinin sınırlamaları gidermek ve insan vücudunun karmaşık hareketlerini daha etkili modellemek amacıyla AGMS-GCN ve FCSA-GCN olmak üzere iki farklı mimari yapı önerilmiştir.

Yöntem: AGMS-GCN mimarisinde, uzaysal-zamansal çizge evrışımleri ile çıkarılan öznelik haritalarına dikkat mekanizması uygulanarak eyleme özgü eklem ilişkilerini temsil eden dikkat bitişiklik matrisleri oluşturulmaktadır. Bu matrisler, kısa ve uzun menzilli uzaysal-zamansal bağımlılıkları etkili bir şekilde yakalayarak eylemler arasındaki ince farkların modellenmesini sağlar. Ayrıca AGMS-GCN modeli, üç farklı ölçekten elde edilen özellik haritalarını kullanarak çok ölçekli uzaysal-zamansal bağımlılıkları modeller. FCSA-GCN modeli ise çok ölçekli özellik çıkarımı için inceden-kabaya ölçekleme stratejisi ve zamansal öz-dikkat mekanizmasını kullanarak yerel ve küresel uzaysal-zamansal bağımlılıkları etkili bir şekilde modeller. Eğitim sırasında çok ölçekli özellik çıkarma aşamalarının dengeli optimizasyonu dinamik ağırlık ortalaması yaklaşımıyla sağlanmıştır.

Bulgular: Kapsamlı deneyler, AGMS-GCN ve FCSA-GCN modellerinin NTU-60 ve N-UCLA veri kümelerinde en son teknoloji performansı sergilediğini ve NTU-120 veri kümesinde rekabetçi sonuçlar elde ettiğini göstermiştir. AGMS-GCN modeli, 15,8M parametre ile NTU-60 veri kümesinde çapraz denek değerlendirmesinde %93,9 ve çapraz görünüm değerlendirmesinde ise %97,8 doğruluk oranına ulaşmıştır. FCSA-GCN modeli ise 5,9M parametre ile çapraz denek değerlendirmesinde %93,6 ve çapraz görünüm değerlendirmesinde ise %97,5 doğruluk oranı yakalayarak daha hızlı ve verimli bir çözüm sunmaktadır.

Sonuç: AGMS-GCN ve FCSA-GCN modelleri, sabit çizge topolojisi sınırlamalarını ortadan kaldırarak uzun menzilli uzaysal-zamansal bağımlılıkları etkili bir şekilde modellemiştir. FCSA-GCN modeli, düşük hesaplama maliyeti ve yüksek doğruluğu sayesinde kaynak kısıtlı ortamlarda da etkili bir şekilde kullanılabileceğini göstermektedir.

Anahtar Kelimeler: iskelet tabanlı eylem tanıma, çizge evrışımsel ağlar, dikkat mekanizması, zamansal öz-dikkat mekanizması, çok ölçekli yapı, inceden-kabaya yaklaşım.

Haziran 2025, 109 sayfa

ABSTRACT

DOCTORAL DISSERTATION

DEVELOPMENT OF SPATIO-TEMPORAL GRAPH CONVOLUTIONAL NETWORK ARCHITECTURES FOR SKELETON-BASED ACTION RECOGNITION GUIDED BY SELF-ATTENTION MECHANISM

Uğur KILIÇ

Supervisor: Assist. Prof. Dr. Gülşah TÜMÜKLÜ ÖZYER
Co-supervisor: Assist. Prof. Dr. Özge ÖZTİMUR KARADAĞ

Purpose: In recent years, skeleton data has become an important modality for action recognition due to its robustness against environmental changes, computational efficiency, compact structure, and capability to provide anonymity. With the development of deep learning, numerous methods have been proposed for action recognition using skeleton data. In particular, the topological structure of skeleton data makes it suitable for modeling as a graph, leading to increased interest in spatial-temporal graph convolutional networks (ST-GCN). However, structural limitations of ST-GCN models, such as fixed graph topology constraints, inadequate representation of dependencies between distant joints, lack of long-range temporal dependencies, and insufficient representation of coordinated movements, remain significant challenges. In this doctoral study, two different architectures, AGMS-GCN and FCSA-GCN, are proposed to address the limitations of ST-GCN models and to more effectively model the complex movements of the human body.

Method: In the AGMS-GCN architecture, attention adjacency matrices representing action-specific joint relationships are generated by applying an attention mechanism to the feature maps extracted through spatial-temporal graph convolutions. These matrices effectively capture short- and long-range spatial-temporal dependencies, enabling the modeling of subtle differences between actions. The AGMS-GCN model represents multi-scale spatial-temporal dependencies by leveraging feature maps obtained from three different scales. The FCSA-GCN model, on the other hand, employs a fine-to-coarse scaling strategy and a temporal self-attention mechanism for multi-scale feature extraction, allowing for more effective modeling of local and global spatial-temporal dependencies. During training, the balanced optimization of multi-scale feature extraction stages is ensured through a dynamic weight averaging approach.

Findings: Comprehensive experiments have shown that AGMS-GCN and FCSA-GCN models demonstrate state-of-the-art performance on the NTU-60 and N-UCLA datasets and achieve competitive results on the NTU-120 dataset. The AGMS-GCN model, with 15.9M parameters, achieves an accuracy of 93.9% in cross-subject evaluation and 97.8% in cross-view evaluation on the NTU-60 dataset. On the other hand, the FCSA-GCN model offers a faster and more efficient solution with 5.9M parameters, achieving 93.6% accuracy in cross-subject evaluation and 97.5% in cross-view evaluation.

Results: The AGMS-GCN and FCSA-GCN models have effectively addressed fixed graph topology limitations, modeling long-range spatial-temporal dependencies with high accuracy. The FCSA-GCN model demonstrates that it can be effectively used in resource-constrained environments due to its low computational cost and high accuracy.

Keywords: skeleton-based action recognition, graph convolutional networks, attention mechanism, temporal self-attention mechanism, multi-scale structure, fine-to-coarse approach.

June 2025, 109 pages

İÇİNDEKİLER

KABUL VE ONAY TUTANAĞI.....	i
ETİK BİLDİRİM VE İNTİHAL BEYAN FORMU	ii
TEŞEKKÜR	iii
ÖZET	iv
ABSTRACT	v
TABLolar DİZİNİ.....	ix
ŞEKİLLER DİZİNİ	xi
KISALTMALAR VE SİMGELER DİZİNİ.....	xiii
GİRİŞ.....	1
Problem Tanımı ve Araştırma Alanının Önemi	1
Araştırma Problemi ve Zorluklar	4
Sabit çizge topolojisi problemi	4
Uzak mesafeli bağımlılıkların yetersiz temsili.....	4
Uzun menzilli bağımlılıkların yetersiz temsili	4
Koordineli hareketlerin yetersiz temsili	5
Tezin Amacı ve Katkıları	5
Tezin Organizasyonu	8
KURAMSAL TEMELLER.....	10
Eylem Tanımadaki Kullanılan Veri Türleri	10
İskelet Tabanlı Eylem Tanıma Yöntemleri	13
El işi tabanlı eylem tanıma yöntemleri.....	13
CNN tabanlı eylem tanıma yöntemleri	14
RNN tabanlı eylem tanıma yöntemleri	16
GCN tabanlı eylem tanıma yöntemleri	18
Transformer tabanlı eylem tanıma yöntemleri	22
MATERYAL VE METOT.....	25
Veri Kümeleri.....	25
NTU-RGB+D 60 veri kümesi	25
NTU-RGB+D 120 veri kümesi	26
Northwestern-UCLA veri kümesi	28
Veri Ön işleme	29

Ötelemeden bağımsızlık.....	29
Kamera bakış açısından bağımsızlık.....	29
Ölçekten bağımsızlık.....	30
Çoklu Modalite Temsili	31
Eklem modalitesi.....	31
Kemik modalitesi	31
Hareket modalitesi	31
Uzaysal-Zamansal Çizge Yapısı	32
Uzaysal kenarlar.....	32
Zamansal kenarlar	33
Bitişiklik matrisleri	33
Çok Ölçekli Çizge Yapıları.....	35
Uzaysal-Zamansal Çizge Evrışimsel Ağlar	37
Uzaysal çizge evrışimi	37
Zamansal evrışim	39
AGMS-GCN Mimarisi.....	39
Mimari yapıya genel bakış	39
Çizge özellik çıkarıcı modülü	43
Dikkat bitişiklik matrisi üretici modülü	43
Sınıflandırma modülü	45
DeneySEL ayarlar.....	45
FCSA-GCN Mimarisi	46
Mimari yapıya genel bakış	46
İnceden-kabaya özellik çıkarma stratejisi	49
Zamansal öz-dikkat	50
Toplam kayıp değerlerinin DWA ile optimizasyonu	53
DeneySEL ayarlar.....	56
ARAŞTIRMA BULGULARI VE TARTIŞMA.....	57
Değerlendirme Protokolleri.....	57
AGMS-GCN Mimarisinin Deney Sonuçları.....	57
Dikkat bitişiklik matrislerinin görselleştirilmesi.....	58
Çoklu modalite öğrenmenin doğruluk üzerindeki etkisi.....	59
Çok ölçekli yaklaşımın doğruluk üzerindeki etkisi.....	59
Öğrenme sürecinin görsel analizi.....	60
Mimarinin sınıf bazlı değerlendirmesi	62

Son teknoloji yöntemlerle karşılaştırmalı doğruluk analizi	65
FCSA-GCN Mimarisinin Deney Sonuçları	67
İnceden-kabaya yaklaşım ile farklı ölçeklerde doğruluk analizi.....	67
Toplam kayıp değerinin optimizasyonun doğruluk üzerine etkisi.....	69
Zamansal öz-dikkat mekanizmasının doğruluk üzerine etkisi	70
Öğrenme sürecinin görsel analizi.....	72
Mimarinin sınıf bazlı değerlendirmesi	73
Son teknoloji yöntemlerle karşılaştırmalı doğruluk analiz	76
AGMS-GCN ve FCSA-GCN Mimarilerinin Karmaşıklık Analizi.....	78
SONUÇLAR ve ÖNERİLER	81
KAYNAKÇA	85
ÖZGEÇMİŞ.....	93

TABLÖLAR DİZİNİ

Tablo 1. Eylem Tanımda Kullanılan Veri Türlerinin Avantajları / Dezavantajları.	11
Tablo 2. NTU-60 ve NTU-120 Veri Kümesinde Yer Alan Eylem Sınıfları ve Kategorileri ..	26
Tablo 3. Kullanılan Veri Kümelerinin Özellikleri	28
Tablo 4. Değerlendirme Protokollerine Dayalı Olarak Veri Kümelerinin Eğitim ve Test Bölümlerine Ayrılması	57
Tablo 5. AGMS-GCN Mimarisinin NTU-60 Veri Kümesinde Farklı Modalitelerde Elde Ettiği Doğruluk (%) Oranlarının Karşılaştırılması	59
Tablo 6. AGMS-GCN Mimarisinin NTU-60 Veri Kümesinde Farklı Ölçekli Çizge Yapılarıyla CS Protokolünde Elde Ettiği Doğruluk Oranları (%).....	60
Tablo 7. AGMS-GCN Mimarisinin NTU-60 Veri Kümesinde Farklı Ölçekli Çizge Yapılarıyla CV Protokolünde Elde Ettiği Doğruluk Oranları (%)	60
Tablo 8. AGMS-GCN Mimarisinin NTU-60 ve NTU-120 Veri Kümeleri Üzerinde Son Teknoloji Yöntemlerle Top-1 Doğruluk (%) Analizi.....	65
Tablo 9. AGMS-GCN Mimarisinin N-UCLA Veri Kümesi Üzerinde Son Teknoloji Yöntemlerle Top-1 Doğruluk (%) Analizi	67
Tablo 10. FCSSA-GCN Mimarisinin NTU-60 Veri Kümesinde CS Protokolüne Göre, Farklı Ölçek ve Modalitelerde İnceden-Kabaya Yaklaşımıyla Elde Ettiği Doğruluk Oranları (%)	68
Tablo 11. FCSSA-GCN Mimarisinin NTU-60 Veri Kümesinde CV Protokolüne Göre, Farklı Ölçek ve Modalitelerde İnceden-Kabaya Yaklaşımıyla Elde Ettiği Doğruluk Oranları (%)	68
Tablo 12. FCSSA-GCN Mimarisinin NTU-60 Veri Kümesinde İnceden-Kabaya ve DWA Yaklaşımlarıyla, Farklı Ölçek ve Modalitelerde CS Protokolüne Göre Elde Ettiği Doğruluk Oranları (%)	69
Tablo 13. FCSSA-GCN Mimarisinin NTU-60 Veri Kümesinde İnceden-Kabaya ve DWA Yaklaşımlarıyla, Farklı Ölçek ve Modalitelerde CV Protokolüne Göre Elde Ettiği Doğruluk Oranları (%)	69
Tablo 14. FCSSA-GCN Mimarisinin NTU-60 Veri Kümesinde Farklı Modaliteler İçin TSA Olmadan Doğruluk (%) Karşılaştırması	71
Tablo 15. FCSSA-GCN Mimarisinin NTU-60 Veri Kümesinde Farklı Modaliteler İçin TSA ile Doğruluk (%) Karşılaştırması.....	71

Tablo 16. FCSA-GCN Mimarisinin NTU-60 ve NTU-120 Veri Kümeleri Üzerinde Son Teknoloji Yöntemlerle Top-1 Doğruluk (%) Analizi	77
Tablo 17. FCSA-GCN Mimarisinin N-UCLA Veri Kümesi Üzerinde Son Teknoloji Yöntemlerle Top-1 Doğruluk (%) Analizi.....	78
Tablo 18. Önerilen Yöntemlerin NTU-60 Veri Kümesinde Son Teknoloji Yöntemlerle Doğruluk ve Karmaşıklık Karşılaştırması.....	79



ŞEKİLLER DİZİNİ

Şekil 1. Eylem tanıma sistemlerinin uygulama alanları	1
Şekil 2. Eylem tanıma sistemlerinde kullanılan görsel ve görsel olmayan veri türleri	10
Şekil 3. (a) İskeletleştirilmiş çerçeve örnekleri (b) CSI temsili (c) CSI'ya karşılık gelen histogram	13
Şekil 4. Bir eylemin iskelet dizisinin Lie gruplar uzayında eğri olarak gösterimi	14
Şekil 5. CNN tabanlı insan eylemi tanıma	15
Şekil 6. RNN tabanlı insan eylemi tanıma	17
Şekil 7. GCN tabanlı insan eylemi tanıma	19
Şekil 8. ST-TR modelinin uzaysal-zamansal öz-dikkat yapısı	23
Şekil 9. NTU-RGB+D veri kümesinden örnek kareler	25
Şekil 10. (a) NTU RGB+D veri kümesine ait 25 vücut eklemının topolojisi (b) Northwestern-UCLA veri kümesine ait 20 vücut eklemının topolojisi.	26
Şekil 11. NTU-RGB+D veri kümesindeki "su içme" eylemi için bir veri ön işleme örneği. (a) Orijinal iskelet verisi (b) Ön işleme adımlarından sonra iskelet verisi	30
Şekil 12. İskelet dizisinin uzaysal-zamansal çizge gösterimi: (a) uzaysal çizge, (b) uzaysal-zamansal çizge	33
Şekil 13. Sabit olarak kullanılan bitişiklik matrisleri: (a) 0-hop bağlantılar, (b) ve (c) 1-hop atlama mesafesinde olan ve sırasıyla merkez düğüme daha yakın ve daha uzak olan düğümler arasındaki bağlantılar, (d) ve (e) sırasıyla merkez düğüme daha yakın ve daha uzak olan 2-hop atlama mesafesinde olan düğümler arasındaki bağlantıları temsil etmektedir	34
Şekil 14. Vücut-parça ilişkisi modeline dayalı olarak oluşturulan farklı ölçeklerdeki çizge yapıları (G_{25} , G_{11} ve G_6)	36
Şekil 15. ST-GC blok yapısı	37
Şekil 16. Yönlendirilmiş çizgenin normalleştirme işlemi sözde kodu	39
Şekil 17. AGMS-GCN'nin genel mimarisi B_i ile temsil edilen her ST-GC bloğunun iç yapısı, şeklin alt kısmında gösterilmiştir	41
Şekil 18. AAMG Modülünün genel mimarisi	44
Şekil 19. FCSA-GCN'nin genel mimarisi B_i ile temsil edilen her ST-GC bloğunun iç yapısı, şeklin alt kısmında gösterilmiştir	47
Şekil 20. Zamansal öz-dikkat mekanizmasının iskelet çerçeveler boyunca çalışması	50

Şekil 21. Zamansal öz-dikkat mekanizmasının iskelet dizilerine uygulanması	52
Şekil 22. (a) Normalize edilmiş sabit bitişiklik matrisleri, (b) AAMG ₁ modülü tarafından üretilen bitişiklik matrisleri, (c) AAMG ₂ modülü tarafından üretilen bitişiklik matrisleri	58
Şekil 23. AGMS-GCN mimarisinin NTU-60 veri kümesi üzerinde CS protokolü kapsamında konum modalitesine ait eğitim grafikleri	61
Şekil 24. AGMS-GCN mimarisinin NTU-60 veri kümesi üzerinde CV protokolü kapsamında konum modalitesine ait eğitim grafikleri	61
Şekil 25. AGMS-GCN mimarisinin NTU-60 veri kümesinde CS protokolü kapsamında elde ettiği, tüm eylem sınıflarını içeren ve hücre değerleri sınıf bazında yüzdelik (%) olarak normalize edilmiş karışıklık matrisi	63
Şekil 26. AGMS-GCN modelinin NTU-60 veri kümesinde, CS protokolü altında en yüksek doğruluk oranına sahip 10 sınıfın sınıf bazlı tahmin dağılım matrisi	64
Şekil 27. AGMS-GCN modelinin NTU-60 veri kümesinde, CS protokolü altında en düşük doğruluk oranına sahip 10 sınıfın sınıf bazlı tahmin dağılım matrisi	64
Şekil 28. Modalite ve ölçeğe göre CS/CV değerlendirme protokollerinde doğrulukları	69
Şekil 29. TSA ile/olmadan farklı modalitelerde doğruluk.	72
Şekil 30. FCSEA-GCN mimarisinin NTU-60 veri kümesi üzerinde CS protokolü kapsamında eklem modalitesine ait eğitim grafikleri	72
Şekil 31. FCSEA-GCN mimarisinin NTU-60 veri kümesi üzerinde CV protokolü kapsamında eklem modalitesine ait eğitim grafikleri	73
Şekil 32. FCSEA-GCN mimarisinin NTU-60 veri kümesinde CV protokolü kapsamında elde ettiği, tüm eylem sınıflarını içeren ve hücre değerleri sınıf bazında yüzdelik (%) olarak normalize edilmiş karışıklık matrisi	74
Şekil 33. FCSEA-GCN modelinin NTU-60 veri kümesinde, CV protokolü altında en yüksek doğruluk oranına sahip 10 sınıfın sınıf bazlı tahmin dağılım matrisi	75
Şekil 34. FCSEA-GCN modelinin NTU-60 veri kümesinde, CV protokolü altında en düşük doğruluk oranına sahip 10 sınıfın sınıf bazlı tahmin dağılım matrisi	76

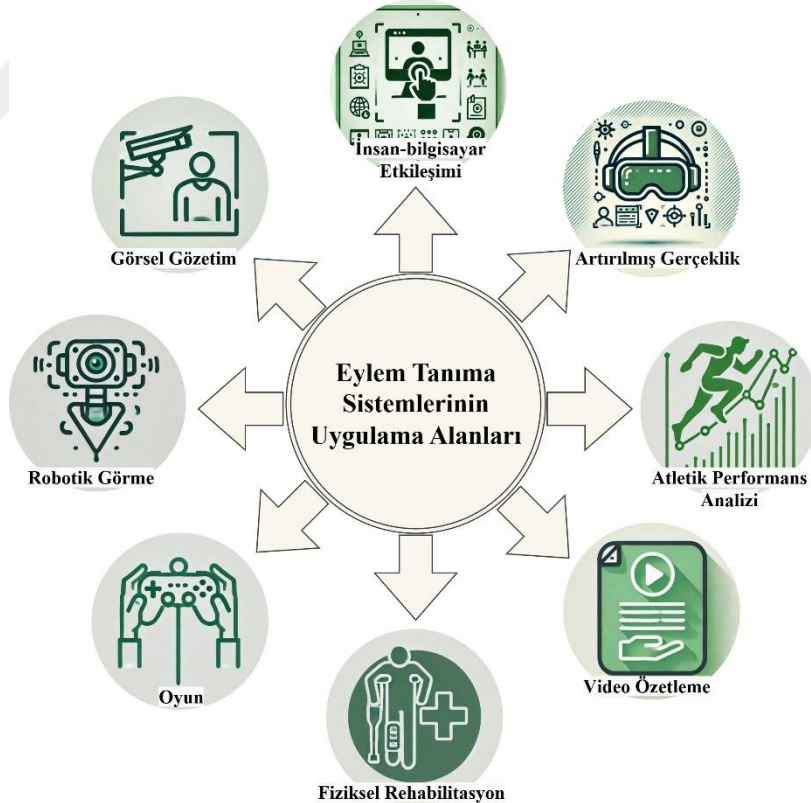
KISALTMALAR VE SİMGELER DİZİNİ

AAAI	: Association for the Advancement of Artificial Intelligence
AAMG	: Attention Adjacency Matrix Generator
ACCV	: Asian Conference on Computer Vision
ACM	: Association for Computing Machinery
AGMS-GCN	: Attention-Guided Multi-Scale Graph Convolutional Network
CE	: Cross Setup
CNN	: Convolutional Neural Network
CS	: Cross Subject
CV	: Cross View
CVPR	: Computer Vision and Pattern Recognition
DWA	: Dynamic Weight Averaging
ECCV	: European Conference on Computer Vision
FCSA-GCN	: Fine-to-Coarse Self-Attention Graph Convolutional Network
FLOPs	: Floating Point Operations Per Second
GAP	: Global Average Pooling
GCN	: Graph Convolutional Neural Networks
GFE	: Graph Feature Extraction
ICCV	: International Conference on Computer Vision
KBS	: Knowledge-Based Systems
LSTM	: Long Short-Term Memory
MHA	: Multi-Head Attention
ReLU	: Rectified Linear Unit
RNN	: Recurrent Neural Network
S-GC	: Spatial Graph Convolutional
SGD	: Stochastic Gradient Descent
ST-GCN	: Spatial-Temporal Graph Convolutional Network
T-GC	: Temporal Graph Convolutional
TIP	: IEEE Transactions on Image Processing
TMM	: IEEE Transactions on Multimedia
TPAMI	: IEEE Transactions on Pattern Analysis and Machine Intelligence
TSA	: Temporal Self-Attention

GİRİŞ

Problem Tanımı ve Araştırma Alanının Önemi

Eylem tanıma, video veya iskelet verisi gibi dinamik veri kaynaklarından bireylerin gerçekleştirdiği eylemleri tespit etmeye ve sınıflandırmaya yönelik bir bilgisayar görüşü ve örüntü tanıma problemidir. Derin öğrenme ve bilgisayar görüşü alanındaki gelişmeler sayesinde, eylem tanıma sistemlerine olan ilgi son yıllarda önemli ölçüde artmıştır (Agahian *et al.*, 2019; Sun *et al.*, 2022). Eylem tanıma sistemleri insan-bilgisayar etkileşimi, artırılmış/sanal gerçeklik (Yang *et al.* 2019; Li *et al.* 2023), video özetlemek (Elharrouss *et al.* 2021), robotik (Terreran *et al.* 2023), atletik performans analizi (Host and Ivašić-Kos 2022; Wu *et al.* 2022), görsel gözetim sistemleri (Singh *et al.* 2019; Verma *et al.* 2021), sağlık hizmetleri (Gao *et al.* 2018; Prati *et al.* 2019), fiziksel rehabilitasyon (Sardari *et al.* 2023), oyun ve akıllı ulaşım (Wan *et al.* 2020; Liang *et al.* 2022) gibi geniş bir yelpazede kritik uygulamalara sahiptir.



Şekil 1. Eylem tanıma sistemlerinin uygulama alanları (Yadav *et al.* 2021; Sun *et al.* 2022)

Güvenlik alanında, eylem tanıma sistemleri anormal davranışların tespiti ve tehlikeli durumlara erken müdahale amacıyla yaygın olarak kullanılmaktadır. Örneğin, kalabalık bir ortamda şüpheli bir hareketi algılayan bir sistem, güvenlik güçlerini uyararak olası bir

tehlikenin önlenmesine katkı sağlayabilir. Bu sayede güvenlik güçleri, toplumsal olayları daha etkili bir şekilde izleyebilir ve acil durumlara daha hızlı müdahale edebilir (Pandey and Kumar, 2024). Sağlık alanında eylem tanıma sistemleri, fiziksel engelli bireylerin kas gücünü yeniden kazanmalarına ve denge yetilerini geliştirmelerine yönelik fizyoterapi egzersizlerinin izlenmesinde önemli bir rol oynamaktadır. Bu sistemler sayesinde, rehabilitasyon sürecinde hastaların gerçekleştirdiği hareketler analiz edilerek tedavi süreçleri daha etkili ve verimli bir şekilde yönetilebilmektedir (Sardari *et al.* 2023). İletişim alanında, işaret dilinin otomatik olarak algılanması, dilsel veya işitsel yetileri bulunmayan bireyler için önemli bir kolaylık sağlamaktadır. Eylem tanıma sistemleri, işaret dilinin doğal hareketlerini algılayarak gerçek zamanlı çeviri yapabilir ve bu sayede iletişim engellerinin aşılmasına yardımcı olabilir (Rastgoo *et al.* 2021; Song *et al.* 2025). Robotik alanında, insan hareketlerinin robotlar tarafından algılanması işbirlikçi robotik sistemlerinin geliştirilmesini sağlar. Bu durum, üretim süreçlerinde insan-robot etkileşimini güçlendirerek iş gücünü azaltabilir ve üretim verimliliğini artırabilir (Rodomagoulakis *et al.* 2016; Terreran *et al.* 2023). Üretim alanında, fabrikalarda çalışanların hareketlerinin izlenmesi, olası hataların tespit edilmesi ve iş akışının optimize edilmesi mümkün hale gelir (Jiao *et al.* 2020). Artırılmış ve sanal gerçeklik uygulamalarında ise eylem tanıma sistemleri, sanal ortamlardaki insan hareketlerini doğru bir şekilde algılayarak avatarların doğal ve gerçekçi hareket etmesini sağlar. Bu durum, sanal gerçeklik deneyimlerinin daha etkileyici ve gerçekçi olmasını mümkün kılar (Li *et al.* 2023). Eylem tanıma, farklı alanlardaki bu geniş uygulama potansiyeli ile yalnızca mevcut sistemlerin performansını artırmakla kalmayıp, gelecekte insan-makine etkileşimini daha doğal ve etkili hale getirecek çözümlere yön verme potansiyeli bulunmaktadır. Özellikle yapay zekâ ve derin öğrenme yöntemlerinin bu alana entegrasyonu, eylem tanımanın doğruluğunu ve etkinliğini artırmaktadır.

Eylem tanıma konusunda yapılan mevcut çalışmalarda, RGB çerçeveleri (Kuehne *et al.* 2011; Carreira and Zisserman 2017), optik akışlar (Simonyan and Zisserman 2014; Sevilla-Lara *et al.* 2019), ses dalgaları (Xiao *et al.* 2020), radar (Song *et al.* 2023), Wi-Fi sinyalleri (Wang *et al.* 2015) ve insan iskeletleri (Yan *et al.* 2018; Shahroudy *et al.* 2016) gibi çeşitli veri türleri kullanılmıştır. Bu veri türleri hakkında detaylı bilgi kurumsal temeller bölümünde verilmiştir. Veri türleri arasında insan iskeletleri eylem odaklı doğası, kompaktlığı sensör gürültüsüne karşı sağlamlıkları ve depolamadaki etkinlikleri nedeniyle, geniş çapta araştırılmış ve büyük ilgi görmüştür. İskelet verileri, insan eklem konumlarının 2B veya 3B koordinatlarını barındıran zaman dizileridir. Bu iskelet dizileri, sadece poz bilgisini içerdiği için yalnızca eylem bilgilerini yakalayarak, arka plan ve ışık değişiklikleri gibi faktörlerden etkilenmezler. Aynı zamanda vücut ölçeklerine ve dinamik kamera görüntülerine karşı da dayanıklıdır (Duan *et*

al. 2022). İskelet verileri gizlilik odaklı doğası ile etik endişeleri en aza indirir ve özellikle sağlık ve güvenlik gibi hassas alanlarda geniş uygulama potansiyeli sunar. Tüm bu özellikleriyle iskelet tabanlı yöntemler, eylem tanıma alanında güçlü bir çözüm olarak öne çıkmaktadır. İskelet verileri, RGB videolara (Sun *et al.* 2019; Cao *et al.* 2017, 2019) veya derinlik haritalarına poz tahmin algoritmaları uygulanarak veya hareket yakalama sistemleriyle elde edilebilir.

İskelet verilerinin birçok avantajına rağmen iskelet tabanlı eylem tanıma yöntemleri birkaç önemli zorlukla karşı karşıyadır. İnsan eylemlerinin sürekli değişen ve karmaşık yapısı, yöntemlerin eklemler arasındaki hem uzaysal hem de zamansal bağımlılıkları etkili bir şekilde temsil etmesini gerektirir. İskelet tabanlı eylem tanıma için geleneksel el işi özellik çıkarımı yapan yöntemler döndürme ve öteleme gibi bir dizi geometrik işlemle iskeletlerin özellikleri manuel olarak çıkarılmaktadır. Fakat bu yöntemler insan hareketlerinin karmaşık uzaysal-zamansal bağımlılıklarını etkili bir şekilde yakalayamadığından sınırlı kalmaktadır.

Son zamanlarda geliştirilen evrişimli sinir ağları (Convolutional Neural Network (CNN)), tekrarlayan sinir ağları (Recurrent Neural Network (RNN)), çizge evrişimli sinir ağları (Graph Convolutional Neural Network (GCN)) ve Transformer tabanlı yöntemler iskelet verilerindeki uzaysal ve zamansal ilişkileri daha etkili bir şekilde modelleyerek bu zorlukların üstesinden gelmeyi amaçlamaktadır. Geleneksel derin öğrenmeye dayalı yöntemler (CNN, RNN) iskeleti eklem koordinat vektörleri dizisi veya sözde görüntü olarak manuel olarak yapılandırır. Bu yapılandırılmış veri CNN veya RNN girdi olarak verilir ve eylem hakkında tahminde bulunmaları beklenir. İskelet verilerini bir vektör dizisi veya sözde görüntü olarak temsil etmek, ilişkili eklemler arasındaki bağımlılığı tam olarak ifade etmez. Bu noktada, GCN'ler insan vücudunun doğal uzaysal yapısını modellemek için etkili bir çözüm sunmaktadır. İnsan vücudu, eklemlerden (düğümler) ve eklemler arasındaki bağlantılardan (kenarlar) oluşan öklid dışı uzayda bir çizge yapısında temsil edilebilmektedir. Bu yapı, ilişkili eklemler arasındaki doğrudan ve dolaylı bağımlılıkların modellenmesini sağlar. Bu nedenle, çizge yapıları eklemler arasındaki ilişkileri doğru bir şekilde modellemek ve zaman içinde hareketleri etkili bir şekilde izlemek için daha uygun hale gelir. İskelet tabanlı eylem tanımayı gerçekleştirmenin en yaygın ve popüler yöntemi, günümüzde iskelet verilerinin öklid dışı doğasını modellemedeki güçlü yetenekleri nedeniyle çizge sinir ağları ve özellikle GCN haline gelmiştir. İskelet tabanlı eylem tanımada GCN'leri ilk olarak kullan çalışma Yan vd. (2018) tarafından yapılmıştır. Uzaysal-zamansal çizge evrişim ağları (Spatial-Temporal Graph Convolutional Networks (ST-GCN)) olarak adlandırılmaktadırlar. ST-GCN mimarileri, iskelet

tabanlı eylem tanıma sistemlerinin performansını önemli ölçüde iyileştirmiş ve insan eylemlerini daha ayrıntılı bir şekilde analiz etme yeteneği sağlamıştır (Yan *et al.* 2018).

Araştırma Problemi ve Zorluklar

ST-GCN, insan vücudundaki eklemler arasındaki doğal bağlantılara dayalı bir uzaysal çizge yapısı oluşturur ve zamansal boyut boyunca ardışık çerçevelerdeki karşılık gelen eklemleri zamansal kenarlarla birleştirir. Daha sonra çizge evrişim katmanları, uzay ve zaman boyunca eylem desenlerini keşfetmek için oluşturulmuş çizgeye kademli olarak uygulanmaktadır. İskelet tabanlı eylem tanıma alanında önemli bir başlangıç noktası olarak kabul edilen ST-GCN modellerinin bazı yapısal sınırlamaları bulunmaktadır (Shi *et al.* 2019b; Plizzari *et al.* 2021).

Sabit çizge topolojisi problemi

ST-GCN modellerinde, insan vücudunu temsil eden çizgenin topolojisi tüm katmanlar ve tüm eylemler için sabitlenmiştir. Bu durum zengin temsillerin ve eylemlere özgü eklemler arasındaki ilişkiyi dikkate alan özniteliklerin çıkarılmasını engellemektedir (Shi *et al.* 2019b; Shi *et al.* 2020; Chen *et al.* 2021).

Uzak mesafeli bağımlılıkların yetersiz temsili

ST-GCN mimarileri uzaysal-zamansal çizgeler üzerinde CNN yapılarında olduğu gibi evrişim çekirdek boyutu tarafından kısıtlanan yerel bir komşulukta standart 2B evrişim işlemi gerçekleştirmektedir. Bu evrişim çekirdek boyutu uzaysal olarak uzak mesafedeki eklemler arasındaki bağımlılıkları yakalamaya engel olmaktadır. Örneğin, sol ve sağ eller arasındaki korelasyonlar, "alkışlama" gibi eylemlerle ilgili olsa bile eksik değerlendirilmektedir. Yani ST-GCN mimarilerinin uzak mesafedeki eklemler arasındaki bağımlılıkları yakalama yeteneği zayıftır. İskelet tabanlı eylem tanıma söz konusu olduğunda, uzak mesafeli bağlamsal bilginin önemi merkezidir. Birçok çalışma, uzak mesafedeki eklemler arasındaki derin anlamsal bilgi elde etmek için hiyerarşik yapıya sahip ağlar (Lee *et al.* 2023) tasarlamıştır ancak sorun tam olarak çözülememiştir. Bu nedenle, uzak mesafeli eklemler arasındaki bağımlılıkları yakalayabilen güçlü uzaysal-zamansal öznitelik geliştirici mimarilerin daha fazla tasarlanması gerekmektedir (Plizzari *et al.* 2021; Wu *et al.* 2024).

Uzun menzilli bağımlılıkların yetersiz temsili

ST-GCN yöntemlerinin bir diğer sınırlaması, zaman içinde hareketlerin karmaşıklığını modellemede önemli bir dezavantaj oluşturan sabit zamansal evrişim çekirdeklerine güvenmeleridir. Örneğin, uzun bir eylem dizisinde başlangıç ve bitiş kareleri arasındaki ilişkiler

kritik olabilir, ancak sabit boyutlu çekirdekler bu tür uzun menzilli bağımlılıkları yakalamada başarısızlığa neden olmaktadır (Chen *et al.* 2021; Plizzari *et al.* 2021).

Koordineli hareketlerin yetersiz temsili

İnsan eylemleri genellikle birden fazla vücut parçasının eş zamanlı ve koordineli hareketini içerir. Ancak, ST-GCN'nin sabit topolojisi, farklı vücut parçaları arasındaki uzaysal koordinasyonu doğru bir şekilde yakalamada sınırlıdır. Bu yapısal sınırlamalar, modelin hareketin tüm dinamiklerini ve karmaşıklığını yeterli şekilde yakalamasını engeller. Örneğin, bir nesneyi kaldırmak gibi basit bir hareket bile kollar, omuzlar, sırt ve bacaklarda bulunan eklem grupları arasında koordineli bir iş birliğini içermesine rağmen bu eklem grupları arasındaki bağımlılıklar, ST-GCN modelleri tarafından dikkate alınmaz. Daha sağlam iskelet tabanlı eylem tanıma yöntemleri tasarlamak için bu koordinasyon dikkate alınmalıdır (Feng *et al.* 2021).

Yukarıda anlatılan sınırlamalar, ST-GCN'nin insan eylemlerinin dinamik ve karmaşık doğasını tam olarak yakalama yeteneğini engeller ve böylece farklı eylemler arasında etkili bir şekilde ayırım yapma kapasitesini kısıtlar. Bu sınırlamaların üstesinden gelmek amacıyla son yıllarda, dinamik çizge yapılarına sahip çeşitli yaklaşımlar önerilmiştir (Shi *et al.* 2019b, 2020; Xie *et al.* 2021). Bu yaklaşımlar, farklı hareket tiplerindeki dinamik eklem ilişkilerini daha etkili bir şekilde modellemeyi amaçlasa da ST-GCN yöntemlerinin sınırlamalarını tamamen aşamamıştır. Ek olarak, dinamik çizge yapılarının kullanımı, modellerin hesaplama maliyetini önemli ölçüde artırarak uygulanabilirliğini sınırlar. Bu nedenle, iskelet tabanlı eylem tanıma alanındaki mevcut zorlukların üstesinden gelmek ve bu yöntemlerin tüm potansiyelini ortaya çıkarmak için daha fazla araştırma yapılması gerekmektedir. Yeni yöntemlerin uzaysal-zamansal bağımlılıkları daha etkili bir şekilde modellemesi yanında hesaplama maliyetlerini azaltarak daha verimli çözümler sunması beklenmektedir.

Tezin Amacı ve Katkıları

İskelet tabanlı eylem tanıma yöntemlerinin performansını iyileştirmek için araştırmacıların ortak fikri, çizge özellik gösterimlerinin eklemler arasındaki uzaysal-zamansal bağımlılığı doğru bir şekilde yansıtmayı sağlamaktır. Bu sebeple GCN tabanlı eylem tanıma yöntemlerinin performansını iyileştirmek için aşağıdaki araştırma soruları belirlenmiştir.

- 1) Uzak eklemler arasındaki uzaysal bağımlılıkların etkili bir biçimde temsil edilmesi nasıl sağlanabilir?
- 2) Modelin daha bilgilendirici zamansal özelliklere odaklanabilmesi için uzun menzilli zamansal bağımlılıklar nasıl yakalanabilir?

3) Farklı eklem grupları arasındaki eş zamanlı ve koordineli hareketler nasıl daha etkili bir şekilde modellenenebilir?

Bu doktora çalışmasında, yukarıda belirtilen yapısal sınırlamaların üstesinden gelmek ve araştırma sorularına cevap olması için AGMS-GCN (Attention-Guided Multi-Scale Graph Convolutional Networks for Skeleton-Based Action Recognition) ve FCSA-GCN (Fine-to-Coarse Self-Attention Graph Convolutional Network for Skeleton-based Action Recognition) olmak üzere iki farklı ağ mimarisi geliştirilmiştir. İlk olarak dikkat rehberliğinde çok ölçekli bir çizge evrişimsel ağ mimarisi olan AGMS-GCN önerilmiştir. AGMS-GCN mimarisi, uzaysal-zamansal çizge evrişimlerini kullanarak çıkarılan özellik haritalarına bir dikkat mekanizması uygulayarak uzaysal-zamansal bağımlılıkları belirler. Bu süreç, iskelet çizge yapısına yeni çizgelerin eklenmesine olanak tanır. Veri odaklı çizge oluşturma yöntemi, modele eylem özgü çizge oluşturma esnekliği sağlar ve farklı eylemlerde daha doğru model temsillerine olanak tanır. AGMS-GCN mimarisi eklem arasındaki ilişkinin önemini öğrenerek, uzak mesafelerdeki eklemler arasındaki etkileşimleri daha doğru bir şekilde modeller. Ek olarak, bir eylemi başarıyla gerçekleştirmek, vücudun farklı bölgelerindeki eklemlerin koordineli hareketini gerektirir. Ancak, literatürdeki mevcut yöntemlerin çoğu bu koordinasyonu göz ardı eder ve tek ölçekli çizge yapıları kullanır. Sonuç olarak, bu yöntemler eylemleri ayırt etmek için gereken yüksek düzeyli anlamsal özellikleri kaçıır. AGMS-GCN mimarisi bu sorunu ele alır ve çok ölçekli yapısı sayesinde daha zengin eylem gösterimleri sağlar. Dikkatle yönlendirilen mekanizmalardan ve çok ölçekli çizge yapılarından yararlanan AGMS-GCN mimarisi, mevcut yöntemlerin sınırlamalarını aşar ve karmaşık insan eylemlerinin daha etkili bir şekilde modellenmesini sağlar. AGMS-GCN mimarisinin temel katkıları aşağıdaki gibi özetlenmiştir.

- AGMS-GCN mimarisi, uzaysal-zamansal çizge evrişimleri tarafından çıkarılan özellik haritaları üzerinde bir dikkat mekanizması kullanarak iskelet çizge yapısına dinamik olarak eylem özgü yeni bitişiklik matrisleri ekler. Bu, kısa ve uzun menzilli uzaysal-zamansal ilişkileri dikkate alan özelliklerin çıkarılmasını sağlar.
- AGMS-GCN'nin veri odaklı yaklaşımı, çizge oluşturmada esneklik sağlar ve farklı veri örneklerine uyum sağlamak için daha fazla genellik sunar.
- Mevcut yöntemlerin genellikle tek ölçekli bir perspektiften hareket ettiği ve bu nedenle yüksek seviyeli anlamsal özellikleri göz ardı ettiği belirtilmiştir. AGMS-GCN, çok ölçekli yapısıyla bu eksikliği giderir ve farklı ölçeklerdeki ilişkileri dikkate alarak daha kapsamlı özelliklerin elde edilmesini sağlar.
- Kapsamlı deneylerimiz, önerilen yöntemin NTU-60 ve N-UCLA veri kümelerinde en son teknoloji performansı sergilediğini ortaya koymuştur. Ayrıca, NTU-120 veri

kümesinde de mevcut yöntemlerle karşılaştırıldığında rekabetçi bir performans göstermiştir.

AGMS-GCN modeli, mevcut yöntemlerin sınırlamalarını ele alarak iskelet tabanlı eylem tanımada önemli ilerlemeler göstermektedir. Dikkatle yönlendirilen bir mekanizma kullanarak, model eklem ilişkilerin önemini dinamik olarak ayarlayarak, doğru eylem tanıma için gereken uzaysal-zamansal bağımlılıkları etkili bir şekilde yakalar. AGMS-GCN modelinin bir diğer güçlü yanı, çok ölçekli mimarisinde yatmaktadır. Farklı ölçeklerdeki üç çizge yapısı ile ayrı ayrı özellik çıkarımı yapılmaktadır. Bu hiyerarşik yöntem, modelin çeşitli veri kümeleri ve eylem türleri arasında genelleştirilebilirliğini artırmaktadır. AGMS-GCN modeli doğruluk açısından literatürdeki yöntemlere göre üstün performans göstermektedir. Fakat dikkat mekanizması ile özellik haritalarından dikkat bitişik matrislerinin oluşturulması ve çok ölçekli yapıların ayrı ayrı işlenmesi yöntemin gerçek zamanlı uygulamalar veya kaynak kısıtlı ortamlar için zorluklar oluşturabilecek hesaplama karmaşıklığı getirmektedir. Bu hesaplama karmaşıklığını azaltmak için çok ölçekli yapıları inceden-kabaya bir yaklaşım kullanarak tek bir akışta birleştirmeye odaklanılarak inceden-kabaya öz-dikkat çizge evrimsel ağ (FCSA-GCN) mimarisi önerilmiştir.

FCSA-GCN mimarisi inceden-kabaya ölçekleme stratejisi, iskelet yapısını farklı ölçeklerde inceleyerek hem yerel ayrıntıları hem de genel hareket kalıplarını etkili bir şekilde öğrenir. Bu birleşik çerçevede, önce yirmi beş eklem noktasını içeren çizge yapısı kullanılarak ince taneli özellik çıkarma işlemini gerçekleştirilmektedir. Bu ayrıntılı yapı, ilk aşamada önemli ve ayrıntılı özellik haritaları elde etmemizi sağlamaktadır. Bu özellik haritaları daha sonra, havuzlama işlemleriyle on bir ve altı eklem noktası içeren çizge yapılarıyla evrim işlemi yapabilecek uygun bir forma dönüştürülmektedir. Daha sonra yirmi beş, on bir ve altı eklem noktası içeren çizge yapıları ile özellik haritaları üzerinde tekrar uzaysal-zamansal evrim işlemi gerçekleştirilerek özellik haritaları birleştirilmektedir. Farklı ölçeklerden elde edilen özellik haritalarının birleştirilmesi hesaplama yükünü optimize ederek hareketin yapısal ve zamansal özelliklerini daha geniş bir perspektifte temsil etmeyi mümkün kılmaktadır. Bu strateji, insan vücudundaki uzaysal ilişkilerin daha anlamlı bir temsilini sağlar ve koordineli hareketleri daha iyi modeller. Geleneksel ST-GCN yöntemleri, yalnızca kısa menzili ilişkileri öğrenebilen sabit boyutlu zamansal evrim çekirdeklerine güvenmektedir. Ancak, FCSA-GCN model mimarisinde, her özellik çıkarma modülünden sonra entegre edilen zamansal öz-dikkat mekanizması, zaman içindeki kritik hareket ayrıntılarını vurgulayarak kareler arasındaki aynı eklem noktasının dinamik değişikliklerini yakalamaktadır. Örneğin, bir "sıçrama" hareketinin başlangıcındaki hazırlık aşaması ile sonundaki iniş aşaması arasındaki bağlantı, zamansal öz-

dikkat mekanizması aracılığıyla güçlü bir şekilde temsil edilmektedir. Zamansal öz-dikkat, bu dinamikleri etkili bir şekilde modelleyerek hem kısa hem de uzun vadeli zamansal hareket kalıplarını yakalamaktadır. Bu yapı öğrenme sürecinin verimliliğini artırır ve mevcut yöntemlerin uzaysal ve zamansal sınırlamalarını aşarak iskelet tabanlı eylem tanıma görevlerinde üstün performans sunmaktadır. Ek olarak, eğitim sırasında çok ölçekli özellik çıkarma aşamalarının dengeli optimizasyonunu sağlamak için dinamik ağırlık ortalaması (Dynamic Weight Averaging (DWA)) yaklaşımı kullanılmıştır. NTU-60, NTU-120 ve N-UCLA veri kümeleri üzerinde yürütülen kapsamlı deneysel çalışmalar önerilen modelin üstünlüğünü açıkça göstermektedir. Önerilen FCSA-GCN modelinin temel katkıları aşağıda özetlenmiştir:

- FCSA-GCN modeli, insan vücudunun topolojik yapısı ve insan eylemlerinin kinematiğine dayanarak oluşturulmuş farklı ölçeklerde üç çizge yapısına sahiptir. İlk özellik çıkarımından sonra model, bu çizge yapılarının nasıl oluşturulduğuna göre özellik haritasında havuzlama işlemleri gerçekleştirir ve farklı ölçeklerdeki çizge yapıları için özellik haritaları oluşturur. Daha sonra, bu farklı ölçeklerden çıkarılan özellikler üzerinde tekrar özellik çıkarımı yapılır ve özellik haritaları birleştirilir. Bu yaklaşım, eylemlerin uzaysal boyutunu inceden-kabaya bir şekilde modeller ve hem vücut eklemlerinin koordineli hareketlerini hem de eklemler arasındaki uzaysal ilişkileri etkili bir şekilde temsil eder.
- Zamansal öz-dikkat mekanizması, eylemlerin zamansal dinamiklerini etkili bir şekilde modeller ve ST-GCN modellerinin ele almadığı uzun vadeli bağımlılıkların yakalanmasını sağlar. Bu mekanizma, eylemin başından sonuna kadar tüm karelerdeki aynı eklem noktaları arasında küresel bir bağlam oluşturarak zamansal bilgileri daha verimli bir şekilde entegre eder.
- DWA yöntemi, farklı ölçeklerden özellik çıkarmada kayıp değerlerinin optimizasyonu sırasında ağırlıkları dinamik olarak ayarlamak için kullanılır. Bu, her ölçeğin katkısının dengelenmesiyle modelin öğrenme sürecinin daha verimli hale getirilmesini sağlar.

Tezin Organizasyonu

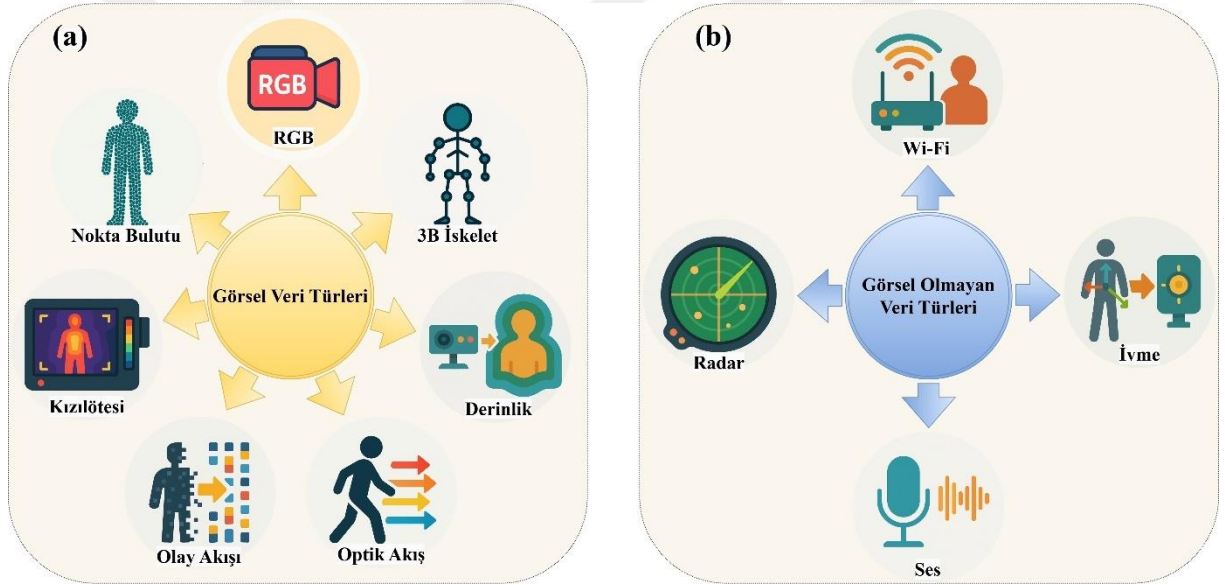
Bu tez, iskelet tabanlı eylem tanıma alanında geliştirilen AGMS-GCN ve FCSA-GCN mimarilerini ele almakta ve bu modellerin farklı veri kümeleri üzerindeki performanslarını analiz etmektedir. Tez toplam yedi ana bölümden oluşmaktadır. Birinci bölüm olan Giriş kısmında, problem tanımı, araştırma alanının önemi ve karşılaşılan temel zorluklar ele alınmaktadır. Ayrıca, tezin amacı, katkıları ve genel yapısı bu bölümde özetlenmektedir. (2)

İkinci bölüm olan Kuramsal Temeller kısmında, literatürde bulunan iskelet tabanlı eylem tanıma yöntemleri ele alınmaktadır. Mevcut yöntemlerin katkıları ve sınırlamaları tartışılmakta, çalışmanın dayandığı teorik temeller ortaya konmaktadır. (3) Üçüncü bölüm olan Materyal ve Metot kısmında, kullanılan veri kümeleri ve değerlendirme protokolleri tanıtılmaktadır. Veri ön işleme adımları ile çoklu modalite temsili açıklanmaktadır. Uzaysal-zamansal çizge yapıları, bitişiklik matrisleri ve çok ölçekli çizge yapıları hakkında ayrıntılı bilgi verilmektedir. Bu bölümde ayrıca önerilen AGMS-GCN ve FCSSA-GCN mimarileri detaylı şekilde açıklanmaktadır. (4) Dördüncü bölüm olan Araştırma Bulguları ve Tartışma kısmında, geliştirilen modellerin deney sonuçları sunulmakta ve son teknoloji yöntemlerle karşılaştırmalı analizler yapılmaktadır. (5) Beşinci bölüm olan Sonuçlar ve Öneriler kısmında, yapılan çalışmaların genel bir değerlendirmesi yapılmakta, elde edilen sonuçlar özetlenmekte ve gelecekte yapılabilecek çalışmalar için öneriler sunulmaktadır.

KURAMSAL TEMELLER

Eylem Tanımada Kullanılan Veri Türleri

Eylem tanıma sistemlerinde kullanılan veri türleri görsel ve görsel olmayan sensörlerden elde edilen veriler olmak üzere iki ana kategoriye ayrılmaktadır (Yadav *et al.* 2021; Sun *et al.* 2022). Eylem tanıma yöntemlerinin ilk çalışmalarında çoğunlukla kolay erişim ve ucuz maliyetlerinden dolayı RGB videolar kullanılmıştır. Fakat son yıllarda sensör teknolojisinde yaşanan gelişmelerle birlikte derinlik haritaları, iskelet verileri ve radar gibi farklı veri türleri de eylem tanıma çalışmalarında kullanılmaya başlanmıştır. Eylem tanıma sistemlerinden kullanılan görsel ve görsel olmayan veri türleri Şekil 2’de verilmiştir.



Şekil 2. Eylem tanıma sistemlerinde kullanılan görsel ve görsel olmayan veri türleri (Sun *et al.* 2022)

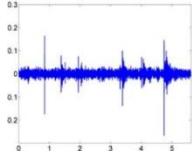
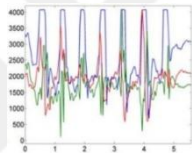
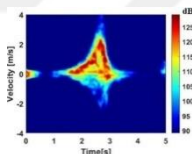
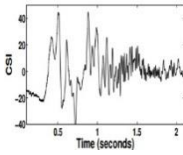
Her veri türünün kendine özgü avantajları ve dezavantajları vardır (Sun *et al.* 2022). Bu avantajlar ve dezavantajlar Tablo 1’de verilmiştir. Örneğin, RGB videoları görsel veriler aracılığıyla eylemleri tespit etmek için yaygın olarak kullanılan bir yöntemdir; ancak değişen bakış açıları, ışık koşulları ve karmaşık arka planlar gibi faktörlerden kolayca etkilenirler (Sun *et al.* 2022). Öte yandan, optik akış yöntemleri hareketin zamansal bileşenlerini yakalamak için etkili araçlardır, ancak genellikle yüksek hesaplama maliyetleri gerektirirler (Piergiovanni and Ryoo 2019). Radar tabanlı yöntemler düşük ışık koşullarında veya fiziksel engellerin olduğu ortamlarda avantajlar sağlar, ancak yakalama cihazlarının nispeten yüksek maliyeti önemli bir dezavantajdır (Li *et al.* 2019). Wi-Fi sinyali tabanlı yöntemler, çevredeki hareketleri algılayarak

alternatif bir yaklaşım sunar; ancak belirli ortamlara duyarlıdır ve gürültülü sinyal koşullarıyla sınırlı olabilir (Wang *et al.* 2017). İskelet verileri ise, insan vücudunun uzaysal yapısına dayanarak oluşturulan çizgeler aracılığıyla her bir eylemin uzaysal ilişkilerini ve zamansal dinamiklerini temsil etme yeteneğiyle öne çıkar. RGB ve optik akış gibi görsel tabanlı yöntemlerin aksine, iskelet verileri doğrudan vücut eklem noktalarını temsil ederek çevresel değişikliklere karşı sağlamlık sağlar. İskelet verileri yalnızca eklem noktalarının koordinatlarını içerdiğinden, yüksek çözünürlüklü görüntü verilerine kıyasla daha az hesaplama kaynağı gerektirir (Duan *et al.* 2022). Ayrıca, mahremiyet endişelerini minimize ederek etik endişeleri en aza indirir ve özellikle sağlık ve güvenlik gibi hassas alanlarda geniş uygulama potansiyeli sunar. Tüm bu özellikleriyle iskelet verileri eylem tanıma sistemleri için güçlü bir çözüm olarak öne çıkmaktadır.

Tablo 1. Eylem Tanımda Kullanılan Veri Türlerinin Avantajları / Dezavantajları.

Veri türü	Örnek	Avantajları / Dezavantajları
RGB	 Saate bakmak (Liu <i>et al.</i> 2017)	RGB görüntüler zengin görsel görünüm bilgisi sunar, kolayca elde edilebilir ve geniş uygulama alanlarına sahiptir. Ancak, bakış açısına, arka plana ve ışık değişimlerine karşı hassastır (Sun <i>et al.</i> 2022).
3B İskelet	 Saate bakmak (Liu <i>et al.</i> 2017)	3B iskelet verisi, özne duruşuna ilişkin yapısal bilgi sağlar. Arka plan, görünüm ve ışık değişimlerine karşı dirençlidir. Ancak görünüm bilgisi sunmaz, detaylı şekil bilgisi eksiktir (Duan <i>et al.</i> 2022).
Derinlik	 Saate bakmak (Liu <i>et al.</i> 2017)	Derinlik verisi, 3B yapısal bilgi ve geometrik şekil bilgisi sunar. Ancak renk ve doku bilgisi sunmaz ve sınırlı çalışma mesafesine sahiptir (Sun <i>et al.</i> 2022).
Optik Akış	 Yürümek (Hoshino and Niimura 2019)	Optik akış, nesnelerin hareket yönü ve hızını kareler arasındaki piksel değişimlerine dayanarak temsil eder; bu sayede hareket bilgisi açısından zengin bir özellik sunar. Ancak, düşük dokuya sahip bölgelerde hassasiyet kaybı yaşanabilir ve hesaplama maliyeti yüksektir (Piergiovanni and Ryoo 2019).
Olay Akışı	 Yumruk atmak (Calabrese <i>et al.</i> 2019)	Olay tabanlı kameralar fazla görsel tekrarları önler, yüksek dinamik aralık sunar ve hareket bulanıklığı içermez. Ancak eş zamansız veri üretir, uzay-zamansal olarak seyrek ve pahalı donanım gerektirir (Calabrese <i>et al.</i> 2019; Sun <i>et al.</i> 2022).

Tablo 1. (devamı)

Kızılötesi	 Saate bakmak (Liu <i>et al.</i> 2017)	Kızılötesi veriler karanlık ortamlarda çalışabilirlik sunar. Ancak renk ve doku bilgisi içermez ve çevresel değişimlere karşı hassastır (Sun <i>et al.</i> 2022).
Nokta Bulutu	 Kazmak (Cheng and Chung 2016)	Nokta bulutları 3B bilgi ve geometrik şekil sunar, görünüm değişimine karşı dayanıklıdır. Ancak renk ve doku bilgisi sunmaz, hesaplama karmaşıklığı yüksektir (Sun <i>et al.</i> 2022).
Ses	 Zıplama ses dalgası (Ofli <i>et al.</i> 2013)	Ses verisi, eylemleri zamansal olarak kolayca konumlandırmayı sağlar. Ancak görünüm bilgisi içermez (Ofli <i>et al.</i> 2022).
İvme	 Zıplamanın ivme ölçümleri (Ofli <i>et al.</i> 2013)	İvme verisi ince ayrıntılı eylem tanıma için kullanılabilir, gizliliği koruyucudur ve düşük maliyetlidir. Ancak görünüm bilgisi içermez ve sensör kişinin üzerinde olmalıdır (Ofli <i>et al.</i> 2013).
Radar	 Düşmek (Li <i>et al.</i> 2023)	Radar, duvar arkası izleme gibi uygulamalar için uygundur, ışık ve hava koşullarından etkilenmez ve gizliliği korur. Ancak görünüm bilgisi sunmaz ve cihazlar pahalıdır (Li <i>et al.</i> 2019).
Wi-Fi	 Düşme için CSI dalga formu (Wang <i>et al.</i> 2015, 2017)	Wi-Fi verisi basit ve kullanışlıdır, gizliliği korur ve düşük maliyetlidir. Ancak görünüm bilgisi sunmaz ve gürültülü olabilir (Wang <i>et al.</i> 2015, 2017).

Görsel ve görsel olmayan verilerin bir arada kullanılması, eylem tanıma sistemlerinin doğruluğunu artırmakta etkili olsa da bu durum veri füzyonunun karmaşıklığını ve hesaplama yükünü artırmaktadır. Farklı veri kaynaklarının senkronizasyonu, veri boyutlarının uyumsuzluğu ve gürültü seviyelerindeki farklılıklar, sistemin performansını etkileyebilecek önemli zorluklar arasında yer alır. Ayrıca, görsel ve görsel olmayan verilerin farklı çözünürlüklerde ve örnekleme hızlarında olması, zaman uyumsuzluklarına ve veri kaybına yol açabilir. Bu nedenle, farklı veri kaynaklarını uygun bir şekilde hizalamak, öz nitelik çıkarmak

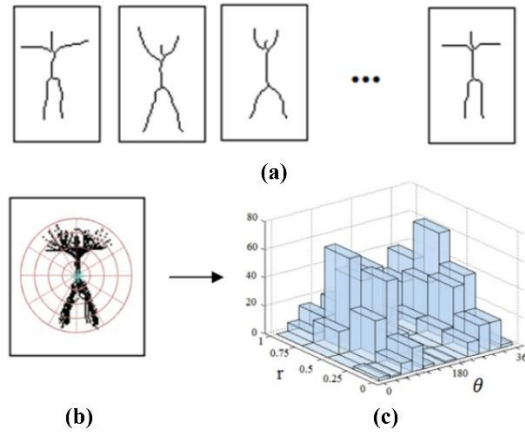
ve ortak bir temsile dönüştürmek için gelişmiş veri füzyonu tekniklerinin kullanılması gerekmektedir (Wang *et al.* 2019; Boulahia *et al.* 2021; Rani *et al.* 2021).

İskelet Tabanlı Eylem Tanıma Yöntemleri

İskelet tabanlı eylem tanıma, eklemlerin 2B/3B koordinatlarını kullanarak insan hareketlerini etkili bir şekilde temsil etme kabiliyeti sayesinde eylem tanıma alanında önemli bir araştırma konusu haline gelmiştir. İskelet tabanlı eylem tanıma konusundaki ilk çalışmaların çoğu, eklemler arasındaki göreceli 3B döndürmeleri ve ötelemeleri kullanan el işi özellik çıkarımına dayanmaktadır. Derin öğrenmenin gelişmesiyle birlikte iskelet verilerinin uzaysal ve zamansal dinamiklerini modellemek için sinir ağı modelleri olan CNN, RNN, GCN ve Transformer tabanlı yöntemler iskelet tabanlı eylem tanıma son teknoloji yöntemler haline gelmiştir. Bu bölümde el işi, CNN, RNN, Transformer tabanlı eylem tanıma yöntemleri ile ilgili literatürde bulunan çalışmaların katkıları ve sınırlamaları hakkında ayrıntılı bir analiz sunulmaktadır.

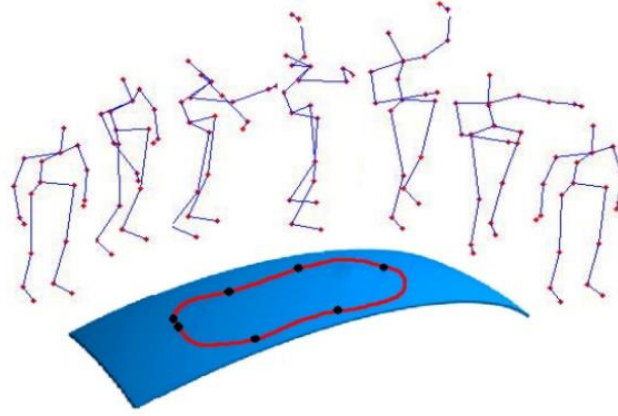
El işi tabanlı eylem tanıma yöntemleri

İskelet tabanlı eylem tanıma için geleneksel yöntemlerde döndürme ve öteleme gibi bir dizi geometrik işlemle iskeletlerin özellikleri manuel olarak çıkarılmaktadır. Ziaeeefard and Ebrahimnezhad (2010) yaptıkları çalışmada tüm iskelet dizilerini kümülatif iskeletleştirilmiş görüntüler (Cumulative Skeletonized Images (CSI)) olarak sıkıştırmış ve ardından her CSI için normalize edilmiş kutupsal histogramı hesaplamışlardır. Yönteme ait görsel Şekil 3'de verilmiştir. Daha sonra iki seviyeli hiyerarşik sınıflandırma kullanarak insan eylemini tanımaya çalışmışlardır. İlk seviyede, histogramın tüm bölümleri ile genel bir sınıflandırma yapılır. İkinci seviyede, daha benzer hareketlerin özel bölümleri kullanılarak tekrar incelenir ve bir nevi ince sınıflandırma yapılmaktadır. İki aşamada da lineer çok sınıflı SVM kullanmışlardır.



Şekil 3. (a) İskeletleştirilmiş çerçeve örnekleri (b) CSI temsili (c) CSI'ya karşılık gelen histogram (Ziaeeefard and Ebrahimnezhad 2010)

Yang and Tian (2012) insan hareketlerini 3B vücut eklem pozisyonlarından tanımak için eklemlerinin konum farklarına dayalı bir özellik türü olan öz-eklemler (EigenJoints) yöntemini önermişlerdir. Öz-eklemler yönteminde statik duruş, hareket ve ofset eylem bilgileri birleştirilmiştir. Öz-eklemlerin sınıflandırılması için Naïve-Bayes en yakın komşu yöntemini kullanmışlardır. Ancak, hesaplama açısından yoğun haritalama süreci nedeniyle bu yöntem gerçek zamanlı olarak çalışması pek mümkün değildir. Vemulapalli *et al.* (2014) yaptıkları çalışmada 3B uzayda geometrik dönüşümleri kullanarak vücut parçaları arasındaki 3B geometrik ilişkileri modelleyen Lie gruplarına dayalı bir iskelet temsili önermişlerdir. Önerilen temsili kullanarak, insan eylemlerini bu Lie grubundaki Şekil 4'te verilen eğriler olarak modellemişlerdir. Lie grubundaki eylem eğrilerinin sınıflandırılmasını kolaylaştırmak için bir vektör uzayı olan Lie cebirine eşlemişlerdir. Daha sonra, dinamik zaman atlama, fourier zamansal piramit temsili ve doğrusal SVM'nin bir kombinasyonunu kullanarak eylem sınıflandırma yapmışlardır.



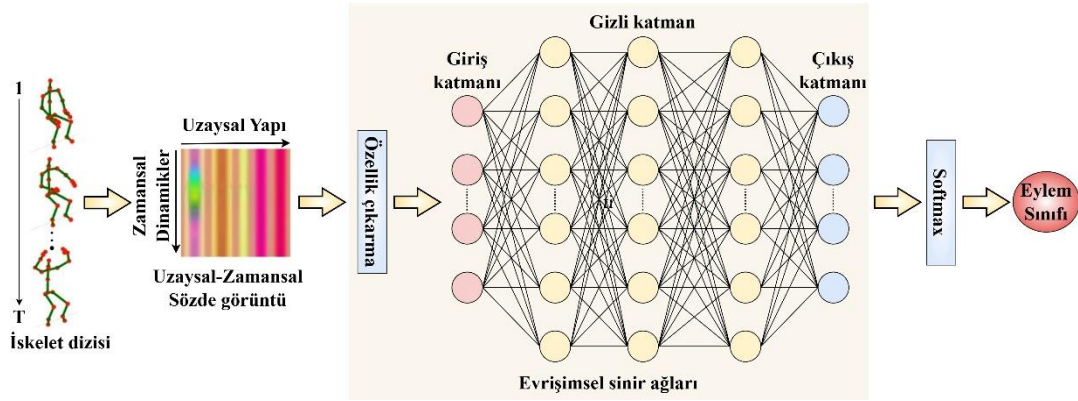
Şekil 4. Bir eylemin iskelet dizisinin Lie gruplar uzayında eğri olarak gösterimi (Vemulapalli *et al.* 2014)

El işi eylem tanıma yöntemleri belli bir oranda başarıya ulaşmış olsa da karmaşık özellik temsili ve yüksek hesaplama maliyeti, el yapımı özellik çıkarımını son derece zorlaştırmaktadır. İnsan iskeletlerinin topolojik yapısı veri kümeleri arasında çeşitlilik gösterdiğinden, el işi eylem tanıma yöntemlerinin genelleştirilebilirliği sınırlı olmaktadır.

CNN tabanlı eylem tanıma yöntemleri

CNN'ler, 2B görüntüler üzerinde yüksek seviyeli anlamsal bilgileri etkili bir şekilde öğrenme konusunda güçlü bir performans sergilemektedir. Bu nedenle iskelet tabanlı eylem tanıma alanında çalışan CNN tabanlı yöntemler, genellikle iskelet verilerini sözde görüntüler olarak temsil etmektedir (Feng *et al.* 2022). Bu sözde görüntülerin oluşturulması için literatürde İskelet haritası (Skeleton map) (Du *et al.* 2015), Eklem uzaklık haritası (Joint distance map) (Li *et al.* 2017), Eklem yörünge haritası (Joint trajectory map) (Wang *et al.* 2018) gibi çeşitli veri

temsil yöntemleri kullanılmaktadır. Daha sonra bu görüntüler üzerine CNN modelleri uygulanarak özellik çıkarımı yapılır ve eylem tanıma işlemi gerçekleştirilir. CNN tabanlı eylem tanıma yöntemlerinin genel diyagramı Şekil 5’de verilmiştir.



Şekil 5. CNN tabanlı insan eylemi tanıma

Du *et al.* (2015), iskelet dizilimini, eklem koordinatlarının kronolojik sırayla düzenlenmesiyle elde edilen bir matris olarak temsil etmişlerdir. Daha sonra bu matrisi görüntüye dönüştürüp normalleştirmiş ve ardından CNN modeline aktararak eylem tanıma gerçekleştirmişlerdir. Ke *et al.* (2017) önerdikleri yöntemde, her iskelet dizilimini üç klipe dönüştürerek, uzaysal-zamansal özellikleri öğrenmek için bir yöntem sunmuşlardır. Bu klipleri, iskeletin silindirik koordinatlarından türetmişlerdir. CNN ve çoklu görev öğrenme ağı kullanarak bu kliplerden uzaysal-zamansal özellikler çıkararak sınıflandırmışlardır. Li *et al.* (2017), eklemlerin 3B koordinatlarını kullanarak 2B görüntü temsilleri oluşturmuşlardır. Daha sonra oluşturulan bu görüntüleri eylemleri tanımak için ön-eğitilmiş VGG-19'a girdi olarak vermişlerdir. Cao *et al.* (2018), iskelet verilerinin düşük boyutlu özellikler olduğunu ve bu özelliklerin kronolojik sırayla bir görüntüye yerleştirilmesiyle orijinal bilginin korunmasının mümkün olduğunu belirtmişlerdir. Vücut eklemlerinin koordinatlarını bir çerçevede iskelet özelliklerine dönüştürürken parça, zincir ve çapraz bazlı sıralamaların performansını sistematik olarak incelemişlerdir. Daha sonra, yığılmış artık bloklardan oluşan bir sınıflandırma ağı oluşturmuş ve daha iyi temsil öğrenimi ile tanıma için doğrusal atlama geçit bağlantıları tasarlamışlardır. Ayrıca, sıralama optimizasyonu amacıyla tam evrişimli bir permütasyon ağı önermişlerdir.

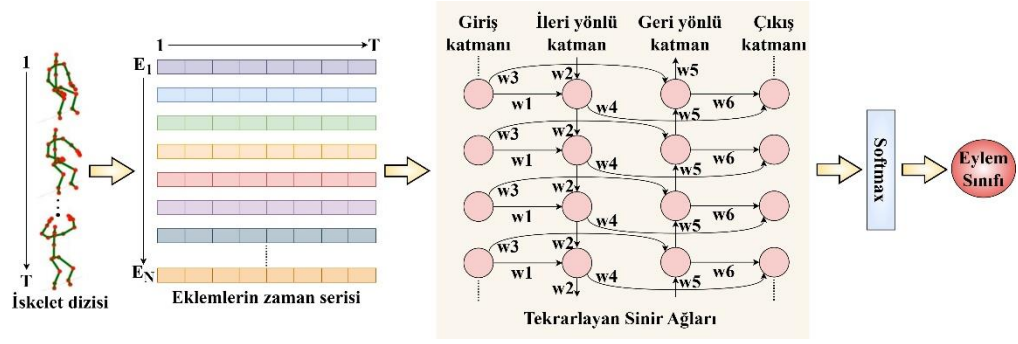
Wang *et al.* (2018), 3B iskelet dizilerindeki uzaysal-zamansal bilgileri üç 2B görüntüyle temsil etmek için bir yöntem önermişlerdir. Bu yöntemde, eklem yörüngeleri ve dinamikleri eklem yörünge haritası adı verilen görüntülerdeki renk dağılımına kodlanarak temsil edilir ve bu görüntüler, insan eylemlerini tanımak için ayrıştırıcı özellikleri öğrenmek amacıyla CNN kullanılarak işlenir. Yöntemin kodlama süreci gereğinden fazla karmaşıktır ve sözde görüntüler, uzaysal özellikleri korusa da eylemlerin hareket bilgisini kaybetmektedirler.

Eylemlerin hareket bilgisini de korumak için Li *et al.* (2019) cebirsel geometri tarafından yönlendirilen, uzaysal özellikleri ve hareketin korelasyonunu aynı anda birleştiren şekil-hareket temsili tasarlamışlardır. Daha sonra derin özellikleri çıkarmak ve birleştirmek için şekil-hareket temsillerine çok akışlı CNN modeli uygulamışlardır. Fakat sıradan 2B CNN modellerini kullanarak, iskelet verileri gibi 3B verilerden uzaysal-zamansal özelliklerini çıkarmak sınırlı öğrenme kapasiteleri nedeniyle zordur. Bu sebeple son yıllarda Duan *et al.* (2022), 3B-CNN tabanlı PoseC3D eylem tanıma yöntemini önermişlerdir. PoseConv3D yönteminde insan iskeletini eklemlerin ve uzuvların uzaysal bağımlılığını yansıtan 2B ısı haritalarının zamansal dizilerinden oluşan 3B ısı haritası hacimleri olarak temsil edilmektedir. Daha sonra bu 3B ısı haritası hacimlerini üzerinde derin özelliklerin çıkarılıp sınıflandırılması için 3B-CNN kullanmışlardır. Kilic vd. (2024), ön-eğitilmiş ResNet mimarisine dayalı SkelResNet modelini önermiştir. SkelResNet modelinin girdi gereksinimlerini karşılamak için yazarlar, iskelet dizilerinden eklem, kemik, eklem hareketi ve kemik hareketi olmak üzere dört farklı sözde görüntü oluşturmuşlardır. Daha sonra, bu sözde görüntüleri SkelResNet modeline girdi olarak vererek modelin özellik çıkarımını gerçekleştirmesini ve eylemleri sınıflandırmasını sağlamışlardır.

Ancak, yapılan tüm bu çalışmalara rağmen CNN tabanlı yöntemler, iskeletin çizge tabanlı yapısını açıkça modelleyememekte, eklemler arasındaki küresel ilişkileri yakalamakta zorlanmaktadır. Ayrıca, iskelet verilerinin sözde görüntülere yerleştirilmesi sırasında iskeletin hareket korelasyonunu hala göz ardı etmektedirler.

RNN tabanlı eylem tanıma yöntemleri

Uzaysal veri analizi için tasarlanmış olan CNN'lerin aksine RNN'ler gizli katmanlarındaki tekrarlayan bağlantılar nedeniyle zamansal verileri analiz etmek için kullanılabilir. RNN'ler uzun vadeli zamansal bağımlılıkları modelleme yetenekleri nedeniyle iskelet tabanlı eylem tanıma da yaygın olarak kullanılmıştır. Bu yöntemlerde, iskelet verileri ilk olarak zaman serilerine dönüştürülür. Daha sonra, zaman serileri üzerinden RNN yapıları kullanılarak eklem noktaları arasındaki zamansal bağımlılıkların modellenmesine odaklanılmaktadır. RNN tabanlı eylem tanıma yöntemlerinin genel diyagramı Şekil 6'da verilmiştir.



Şekil 6. RNN tabanlı insan eylemi tanıma

Du *et al.* (2015) insan iskeletini fiziksel yapısına göre beş vücut parçasına ayıran ve her bir parçayı ayrı alt ağlarla işleyen hiyerarşik bir RNN yapısı önermişlerdir. Daha sonra bu alt ağların çıktı temsilleri, eylemin yüksek düzeyli temsillerini oluşturmak için hiyerarşik olarak birleştirilmektedir. Wang and Wang (2017) tarafından insan iskeletinin hem uzaysal hem de zamansal konfigürasyonları modellemek için yığılmış RNN ve hiyerarşik RNN olmak üzere iki akışlı RNN mimarisi önermişlerdir. Hiyerarşik RNN, insan vücut kinematiğine göre tasarlanmıştır. Fakat ağ yapısı uzaysal alandaki eklemler arasındaki bağımlılıkları modelleme gücü zayıf kalmıştır.

Geleneksel RNN'ler uzun vadeli bağımlılıkları modellemede kaybolan gradyan probleminden müzdariptir. Bu yüzden RNN'ler uzun vadeli zamansal bağımlılıkları etkili bir şekilde modelleme konusunda yetersiz kalmaktadır. Bu nedenle, eylem tanıma alanında kullanılan yöntemlerin çoğu, eklemlerin zamansal bağımlılığını daha iyi modellemek için uzun kısa süreli bellek (Long Short-Term Memory (LSTM)) yapılarını kullanmaktadır. Liu *et al.* (2016), 3B iskelet verilerinin uzaysal-zamansal bağımlılıkları birlikte analiz edebilmek için ağaç yapısı tabanlı tarama yöntemi önermişlerdir. İskelet verilerindeki gürültü ve tıkanma problemlerini ele almak ve sıralı giriş verisinin güvenilirliğini öğrenerek bellekte depolanan uzun vadeli bağlamsal bilgiyi daha etkili şekilde güncellemek için LSTM içinde bir kapı mekanizması tasarlamışlardır. Lee *et al.* (2017), farklı zaman adımı boyutlarına sahip LSTM ağlarının özellikleri iyi modelleyebileceğini belirterek, iskelet tabanlı eylem tanıma için zamansal kayan LSTM (Temporal Sliding LSTM (TS-LSTM)) ağ yapısını önermişlerdir. Önerilen ağ kısa, orta ve uzun vadeli TS-LSTM ağlarını içeren birden fazla bölümden oluşmaktadır. Ancak, TS-LSTM ağ yapısı insan iskeletinin uzaysal konfigürasyonunu yeterince dikkate alamamaktadır. Liu *et al.* (2017), daha bilgilendirici eklemlere seçici olarak odaklanmak ve bunların uzaysal bağımlılıklarını daha da artırmak için iskeletlerin bağlam ilişkisinden yararlanan bir uzaysal-zamansal dikkat LSTM ağı önermişlerdir. Zhang *et al.* (2019), bakış açısı değişikliklerini ele almak için görünüm uyarlama şeması yaklaşımını geliştirmişlerdir. Bu yöntem, eylemler için en uygun bakış açılarını öğrenmek üzere LSTM ve

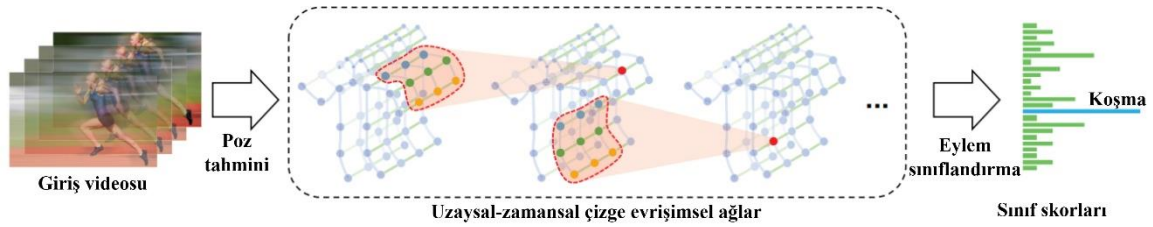
CNN tabanlı iki ağ sunmuştur. Bu ağlar, iskeletleri tutarlı bakış açılarına dönüştürerek eyleme özgü özelliklerin daha iyi öğrenilmesini sağlamayı hedeflemiştir. Nihai sınıflandırma için bu ağların puanlarını birleştirerek daha yüksek performans elde etmişlerdir.

Genel olarak, RNN tabanlı yöntemler zamansal bağımlılıkları modellemede etkili olsalar da iskelet yapısındaki eklemler arasındaki uzaysal konfigürasyonları algılama yeteneğinin sağlanması, bu yöntemler için hala büyük bir zorluk oluşturmaktadır.

GCN tabanlı eylem tanıma yöntemleri

GCN çizge yapılarını girdi olarak alan bir CNN türüdür. Kipf and Welling (2016), çizge fourier dönüşümünü kullanarak frekans alanında çizge evrişimi gerçekleştirmişlerdir. İnsan iskeleti doğal topolojik yapıya sahiptir ve bir çizge ile modellenabilir. Çizgelerin yapısal özelliklerinden yararlanarak, GCN'ler eklemler ve bunların dinamikleri arasındaki karmaşık ilişkileri zaman içinde yakalamak için sağlam bir çerçeve sağlar ve bu da onları iskelet tabanlı eylem tanıma gibi görevler için özellikle uygun hale getirir. Bu fikrin rehberliğinde, birçok çalışma (Yan *et al.* 2018; Shi *et al.* 2019a, 2019b, 2020; Li *et al.* 2019; Shiraki *et al.* 2020; Liu *et al.* 2020; Zhang *et al.* 2020; Cheng *et al.* 2020a; Chen *et al.* 2021; Chen *et al.* 2021; Xie *et al.* 2021; Peng *et al.* 2021; Li *et al.* 2021; Lee *et al.* 2023; Geng *et al.* 2024; Zhou *et al.* 2024; Xie *et al.* 2024; Wu *et al.* 2024; Xu *et al.* 2025) eylem tanıma için çizge yapılarını üzerinde evrişim işlemini gerçekleştiren GCN'ler ile dikkate değer bir performans elde etmiştir.

Eklemler arasındaki uzaysal ve zamansal bağımlılıkları keşfetmek için GCN yapısını ilk kez kullanan yöntem, Yan *et al.* (2018) tarafından geliştirilen uzaysal-zamansal çizge evrişimsel ağ yapısı olan ST-GCN mimarisidir. ST-GCN iskeletlerin uzaysal özelliklerini çıkarmak için çizge evrişim operatörlerinin ve zamansal hareket korelasyonlarını öğrenmek için 1-Boyutlu evrişim operatörlerinin kullanıldığı bir ağ yapısıdır. ST-GCN yapısı eylem tanıma konusunda o zaman kadar geliştirilen yöntemlere göre çok başarılı sonuçlar elde etmiştir ve hızla eylem tanıma konusunda mihenk taşı haline gelmiştir. ST-GCN mimarisinde ilk olarak iskelet verileri üzerinden uzaysal-zamansal çizge oluşturulur. Daha sonra bu çizge yapıları üzerine ST-GCN'nin çoklu katmanları uygulanarak kademeli olarak çizge yapısı üzerinden yüksek seviyeli özellik haritaları oluşturulur. Eylem tanıma konusunda yapılan son çalışmalar uzaysal-zamansal çizgelere veya türevlerine dayanmaktadır. GCN tabanlı eylem tanıma yöntemlerinin genel diyagramı Şekil 7'de verilmiştir.



Şekil 7. GCN tabanlı insan eylemi tanıma (Yan *et al.* 2018)

İskelet verilerindeki hem eklem hem de kemik bilgileri, hareket tanıma görevleri için büyük yardım sağlamaktadır. Ancak, eklem ve kemik bilgileri arasındaki ilişkiyi en iyi şekilde nasıl kullanılacağı hala çözülmesi gereken bir sorundur. Shi *et al.* (2019a), iskelet verilerini doğal insan vücudundaki eklem ve kemikler arasındaki kinematik bağımlılığa dayalı olarak yönlendirilmiş çevrimsiz çizge olarak temsil etmişlerdir. Daha sonra eklem, kemik ve ilişkileri hakkındaki özellikleri çıkarmak ve çıkarılan özelliklere dayalı tahminler yapmak için DGNN (Directed Graph Neural Network) isimli bir çizge sinir ağı tasarlamışlardır. ST-GCN tarafından oluşturulan çizge yerleştirmelerini daha güçlü hale getirmek için Shi *et al.* (2019b), tarafından 2s-AGCN (Two-Stream Adaptive Graph Convolutional Network) geliştirilmiştir. 2s-AGCN, iskelet tabanlı eylem tanıma için hem birinci dereceden (eklemler) hem de ikinci dereceden (kemiklerin uzunlukları ve yönleri) bilgilerini eşzamanlı olarak modelleyen iki akışlı çizge evrişim ağıdır. Aynı zamanda, yazarlar modelin çizgeyi temsil eden bitişiklik matrisinin değerlerini otomatik olarak güncelleyebilmesini sağlamak amacıyla uyarlanabilir çizge stratejisi tasarlamışlardır.

Li *et al.* (2019), eylemsel-yapısal çizge evrişimsel ağ olan AS-GCN'yi (Actional-Structural Graph Convolutional Networks) önermişlerdir. AS-GCN'de eylemsel bağlantılar eyleme özgü gizli bağımlılıkları yakalamak için yapısal bağlantılar ise daha yüksek dereceli bağımlılıkları temsil etmek için tasarlanmıştır. AS-GCN eklemlerin etkileşiminin eylemin önemli bir özelliği olduğunun farkındadır. Fakat bu etkileşimlerin zaman içerisinde değiştiğini dikkate almamışlardır. Eklemlerin öneminin her eyleme göre değiştiği gerçeği ışığında Shiraki *et al.* (2020) her eylem için eklemler arasındaki statik ilişkileri ve her çerçeve için eklemlerin dinamik önemini dikkate alan STA-GCN (Spatial-Temporal Attention Graph Convolutional Networks) mimarisini önermişlerdir. STA-GCN mimarisi her eylem için eklemler arasındaki statik ilişkiyi temsil eden bir dikkat kenarı ve her çerçeve için eklemlerin dinamik önemini temsil eden bir dikkat düğümü oluşturmaktadır. Daha sonra dikkat düğümü özellik haritasına uygulanarak eylem için önemli olan eklemler vurgulanır. Yazarlar özellik haritasını dikkat kenarıyla birleştirerek eklemler arasındaki önemli ilişkileri dikkate alan özellikler elde etmeye çalışmışlardır. Fakat modelin karmaşıklığı ve hesaplama maliyeti yüksektir. Liu *et al.* (2020), çok ölçekli çizge evrişimlerini ayırtıran bir yöntem ve birleşik bir uzaysal-zamansal çizge

evrişim operatörü olan G3D önermişlerdir. Önerilen çok ölçekli birleştirme şeması, farklı komşuluklardaki düğümlerin önemini ayrıştırarak etkili uzun menzilli modellemeyi, G3D modülü ile de yoğun uzay-zaman bağlantıları kullanarak uzun menzilli ve çok ölçekli uzaysal-zamansal bağımlılıkları modellemeyi amaçlamışlardır. Shi *et al.* (2020), iskelet tabanlı eylem tanıma için MS-AAGCN mimarisini tanıtmışlardır. MS-AAGCN, girdi verilerine dayalı çizge yapılarının uyarlanabilir öğrenimine olanak tanır. Model önemli eklemlere, çerçevelere ve özelliklere odaklanmak için uzaysal-zamansal kanal dikkat modülü içermektedir.

Zhang *et al.* (2020), çalışmalarında eylem tanıma için eklemlerin yüksek düzey semantiğini ele almışlardır. İskelet tabanlı hareket tanıma için anlam rehberli sinir ağı olan SGN'yi (Semantics-Guided Neural Network) önermişlerdir. Yazarlar yüksek seviyeli eklem semantiğini (eklem türü ve çerçeve indeksi) ağa açıkça tanıtarak özellik temsil yeteneğini artırmaya çalışmışlardır. GCN'lerin hesaplama maliyetlerini azaltmak için, Cheng *et al.* (2020a) tarafından ağır düzenli çizge evrişimleri kullanmak yerine kaydırma çizge işlemlerini ve hafif noktasal evrişimleri benimseyen ShiftGCN tasarlamışlardır.

Yerel çizge evrişimlerinin uzun menzilli bağımlılıkları yakalamadaki sınırlamalarını ele almak için Chen *et al.* (2021), MST-GCN modelini önermişlerdir. Bu model, yerel çizge evrişimlerini alt çizge işlemlerine ayrıştırarak hiyerarşik bir mimari sunmaktadır. Chen *et al.* (2021), kanal bazlı topoloji iyileştirme yaklaşımıyla çizge evrişimlerini güçlendirmeyi amaçladıkları CTR-GCN (Channel-wise Topology Refinement Graph Convolution) mimarisini önermişlerdir. CTR-GCN'de tüm kanallar için paylaşılan bir genel topoloji belirlenir ve ardından kanal bazlı korelasyonlarla rafine edilmektedir. Xie *et al.* (2021), bitişiklik matrisinde uzaysal konfigürasyonları daha iyi öğrenmek için çizge evrişimine yardımcı olan bir dikkat mekanizması tanıtmışlardır. Ancak, yöntem önemli zamansal özelliklere dikkat etmemektedir. Peng *et al.* (2021), uzaysal-zamansal çizgede önemsiz düğümleri birleştirebilen veya silebilen çizge üçlü havuzlama yöntemini önermişlerdir. Li *et al.* (2021), hem eylem tanıma hem de hareket tahmini görevlerini aynı anda yerine getirmek için simbiyotik GCN mimarisini önermişlerdir. Simbiyotik GCN bir omurga, eylem tanıma kafası ve hareket tahmin kafasından oluşmaktadır. İki kafa birlikte eğitilir ve birbirini geliştirir. Omurga için, uzaysal ve zamansal özellikleri çıkarmak amacıyla çok dallı, çok ölçekli çizge evrişim ağları tasarlamışlardır.

GCN'lerde anlamsal olarak anlamlı kenarlara sahip bitişiklik matrisler oluşturmak eylem tanıma için kritik bir görevdir. Ancak bu tür kenarları çıkarmak zorlu bir problemdir. Bu problemi çözmek için Lee *et al.* (2023) hiyerarşik olarak ayrıştırılmış çizge (HD-Graph) ile hiyerarşik olarak ayrıştırılmış çizge evrişimli ağ olan HD-GCN mimarisini önermişlerdir. Önerilen HD-GCN, yapısal olarak önemli bitişik ve uzak kenarları çıkarmak için her eklem

düğümünü birkaç kümeye ayırıştırır ve bunları, insan iskeletinin aynı anlamsal alanlarındaki bu kenarları içeren HD-Graph oluşturmak için kullanır. Dikkat rehberliğindeki hiyerarşi toplama modülü ile baskın hiyerarşik kenar kümelerini vurgulamayı amaçlamışlardır. Zhou *et al.* (2024), GCN'lerde bitişiklik matrisi model ağırlıklarıyla birlikte optimize edildiğinden, zamanla kemik bağlantı bilgilerinin zayıfladığını ve modelin iskelet topolojisini doğru şekilde temsil etme yeteneğinin azaldığını savunmuşlardır. Bu sorunu çözmek için çizge mesafelerini kullanarak fiziksel topolojiyi koruyan ve kalıcı homoloji analiziyle eyleme özgü topolojik bilgileri dahil eden bir yaklaşım önermişlerdir. Xie *et al.* (2024), GCN tabanlı yöntemlerin genellikle iskelet topolojisini iyileştirmeye odaklanırken, eklem ve kenar türlerini dikkate almadıklarını belirtmişlerdir. Bu eksikliği gidermek amacıyla dinamik anlamsal tabanlı çizge evrişimli ağ yapısı olan DS-GCN'yi (Dynamic Semantic-Based Graph Convolution Network) önermişlerdir. DS-GCN yapısında eklem türü farkındalıklı uyarlanabilir topoloji ve kenar türü farkındalıklı uyarlanabilir topoloji olmak üzere iki anlamsal modül bulunmaktadır. Bu modüller, iskelet topolojisini örtük bir şekilde kodlayarak, hareketlerin anlamsal ilişkilerini daha iyi temsil etmeye çalışmaktadır. Geng *et al.* (2024), ince taneli eylem gösterimlerinin yakalanmasını ve zamansal özellik geliştirmeyi iyileştirmek için hiyerarşik bir çizge evrişim modülü, piramit dikkat birleştirme mekanizması ve çok ölçekli zamansal evrişim modülünü içeren HA-GNN'yi önermişlerdir.

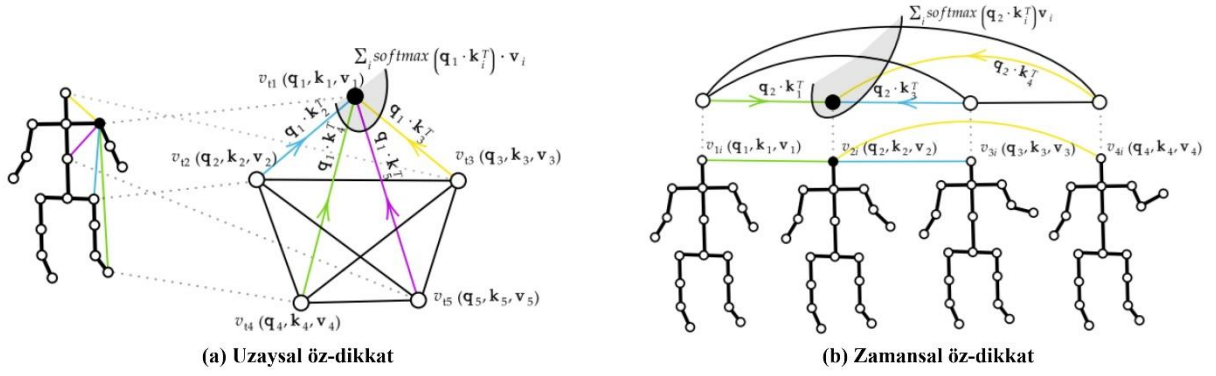
Yukarıdaki dinamik yöntemlerde topoloji modellemesi yalnızca eklem bilgilere dayanmaktadır. Buna karşılık, Xu *et al.* (2025) tarafından geliştirilen bilgi destekli bir dil modeli olan LA-GCN (Knowledge-Assisted Graph Convolutional Network) topolojiyi öğrenmede eylemle ilgili önsel bilgileri dahil etmek için büyük ölçekli dil modeli kullanmaktadır.

GCN tabanlı yöntemler iskelet tabanlı eylem tanımda önemli başarılar elde etmiş olsa da uzak eklem ilişkilerini ve uzun menzilli zamansal bağımlılıkları modelleme yeteneği hala geliştirilmeye ihtiyaç duymaktadır. Anlamsal olarak önemli kenarlara sahip bitişiklik matrislerinin oluşturulması, GCN'lerde eylem tanıma performansını artırmanın anahtarıdır. Fakat, bu tür kenarları çıkarmak oldukça zordur, çünkü iskelet yapısında doğrudan bağlantılı olmayan eklemler arasındaki ilişkileri belirlemek, hareketin bağlamsal dinamiklerini doğru şekilde yakalamayı gerektirir. Geleneksel yöntemler genellikle sabit topolojilere dayanırken, bu tür uzun menzilli bağımlılıkları etkili bir şekilde modelleyebilmek için dinamik ve veri odaklı topoloji öğrenme yaklaşımlarına ihtiyaç duyulmaktadır. Ek olarak, eylemlerin koordineli eklem hareketlerinden kaynaklandığı gerçeği, daha iyi tanıma performansı için uzaysal-zamansal bağımlılıklara dahil edilmesi gerekmektedir.

Transformer tabanlı eylem tanıma yöntemleri

Transformer, Vaswani *et al.* (2017) tarafından doğal dil işleme problemlerinde tekrarlayan sinir ağlarına bir alternatif olarak önerilmiştir. Bu mimari, RNN'lerin çok uzun dizileri işlerken karşılaştığı sorunları aşmak ve kelime kelime gerçekleştirilen cümle işlemenin paralelleştirilmesi için tasarlanmıştır. Transformerlar bir diziye girdi olarak alır ve içerisindeki öz-dikkat mekanizması ile dizideki öğeler ve bunların bağımlılıkları arasındaki ilişkilerin daha iyi temsil edildiği başka diziye dönüştürür. Transformer içindeki çok başlı dikkat mekanizması ise, her bir başın farklı bağımlılıkları öğrenmesini sağlayarak daha zengin temsiller elde edilmesine olanak tanımaktadır. Transformer ve öz-dikkat mekanizması son yıllarda birçok bilgisayar görüşü görevine başarıyla uygulanmıştır (Wang *et al.* 2018; Dosovitskiy *et al.* 2020).

Eylem tanıma performansını iyileştirmek için çizge özellik gösterimlerinin eklemler arasındaki uzaysal-zamansal bağımlılığı daha doğru bir şekilde yansıtması gerekmektedir. Bu bağlamda eklemler arasındaki bağımlılıkların gücünü dinamik olarak ağırlıklandırmak için dikkat mekanizmasını kullanılabilir. Shi *et al.* (2020) iskelet tabanlı eylem-jest tanıma için ayrıştırılmış uzaysal-zamansal dikkat ağı adı verilen bir yaklaşım sunmuşlardır. Bu yöntem öncelikle, önceden tanımlanmış geçiş kurallarına veya çizge topolojilerine ihtiyaç duymadan, iskelet verilerindeki eklemler arasındaki uzaysal-zamansal bağımlılıkları modellemek için dikkat mekanizmalarını kullanır. Fakat yaklaşım uzaysal bağımlılıkları modellemede zayıf kalmaktadır. Cho *et al.* (2020) eylem dizisini zamansal kliplerde bölümlere ayırarak yerleştirmelerin çıkarıldığı bir öz-dikkat ağı olan SAN'nı (Self-Attention Network) önermişlerdir. Uzun vadeli semantik bilgiyi modellemek için yerleştirmelere öz-dikkat uygulamaktadırlar. İskelet eklerimden ziyade yerleştirmelere öz-dikkat uyguladığından, çerçeveler içindeki eklemlerin korelasyonlarını kapsamlı bir şekilde modellemede yetersiz kalmaktadır. Bu problemi çözmek için Plizzari *et al.* (2021) hem çerçeve içi hem de çerçeveler arası çizgedeki düğümlere doğrudan öz-dikkat uygulayarak, iskelet dizisinin hem uzaysal hem de zamansal bağımlılıklarını verimli bir şekilde modellemeyi amaçlayarak iki akışlı uzaysal-zamansal transformatör ağı olan ST-TR'yi (Spatial-Temporal Transformer Network) önermişlerdir. Uzaysal öz-dikkat ve zamansal öz-dikkat her bir düğüm çifti arasındaki korelasyonlarının gücünü temsil eden bir ağırlık hesaplayarak çalışır. ST-TR modelinin uzaysal-zamansal öz-dikkat yapısının görseli Şekil 8'de verilmiştir.



Şekil 8. ST-TR modelinin uzaysal-zamansal öz-dikkat yapısı (Plizzari *et al.* 2021)

ST-TR modelinin uzaysal-zamansal öz-dikkat yapısı her vücut eklemi için ilk olarak bir sorgu (q), bir anahtar (k) ve bir değer vektörü (v) hesaplar. Daha sonra, mevcut eklem (omuz eklemi) sorgusu ile diğer tüm eklemlerin anahtarı arasındaki nokta çarpımı gerçekleştirilir. Böylece her bir eklem çifti arasındaki bağlantı gücünü temsil edecek ağırlık değerleri elde edilir. Elde edilen ağırlık değerleri kullanılarak, her bir eklem kendi korelasyonu ile ölçeklenmesi gerçekleştirilir. Bu ölçekleme işlemi, bir eklem kendi korelasyonu ile orantılı bir ağırlık almasını sağlar. Yani, bir eklem mevcut eklem ile yüksek korelasyona sahipse, daha fazla vurgu alacak ve daha güçlü bir etkiye sahip olacaktır. Belirli eklemler arasındaki örtük bağımlılıkların dikkate alınmasının gereksiz olduğu durumlarda, ST-TR bu eklemler arasındaki korelasyonu gerçekte olduğundan fazla tahmin etmektedir. Örneğin, “Çay içme” ve “El çırpma” eylemlerinde sol ve sağ ayağı temsil eden eklemler her iki eylem sırasında da büyük ölçüde hareketsiz kalmaktadır. Bu eklemler arasındaki ilişki düzeyinin artması, yalnızca hesaplama yükünü yükseltmekle kalmaz, aynı zamanda modelin eylemleri ayırt etmesini zorlaştırarak karışıklığa neden olmaktadır.

Liu *et al.* (2022), çekirdek dikkat uyarlamalı çizge transformer ağı olan KA-AGTN (Kernel Attention Adaptive Graph Transformer Network) mimarisini önermişlerdir. KA-AGTN, eklemler arasındaki yüksek dereceli uzaysal bağımlılıkları, çok başlı dikkat mekanizmasına dayalı çizge transformer operatörü ile modellemektedir. KA-AGTN zamansal çekirdek dikkat bloğu ile zamansal özellikleri kullanarak kanal seviyesinde bir dikkat skoru oluşturur ve bu da zamansal hareket korelasyonunu geliştirmek kullanılmaktadır. Do *et al.* (2024), iskelet tabanlı eylem tanıma için iskelet-zamansal transformer (Skeletal-Temporal Transformer) yöntemini önermişlerdir. Bu yöntem, eklemleri ve çerçeveleri farklı iskelet-zamansal ilişki türlerine göre böler ve her bir bölmede iskelet-zamansal öz-dikkat uygulamaktadır.

Transformerlar uzun menzilli bağımlılıkları yakalamada mükemmel olsa da genellikle insan vücudunun yapısıyla ilgili önceden edinilmiş bilgileri açıkça dahil etmeden verilerden

uzaysal bağımlılıkları öğrenirler. Buna karşılık, insan iskeletinin bileşiminin topolojik yapısı, model optimizasyonu için bir rehber olan insan vücudunun doğal bağlantılarına ilişkin önsel bilgi içermektedir. Bu durum, transformerların GCN tabanlı yöntemlere kıyasla uzaysal özellikleri öğrenmede yetersiz kalmasına neden olabilmektedir.



MATERYAL VE METOT

Veri Kümeleri

NTU-RGB+D 60 (Shahroudy *et al.* 2016), NTU-RGB+D 120 (Liu *et al.* 2019), Northwestern-UCLA (Wang *et al.* 2014) literatürde eylem tanıma konusunda kabul görmüş üç büyük veri kümesidir. Bu veri kümeleri üzerinde sıfırdan eğitim işlemleri yapılarak kapsamlı deneyler sonucunda geliştirilen mimarilerin performansları analiz edilmiştir.

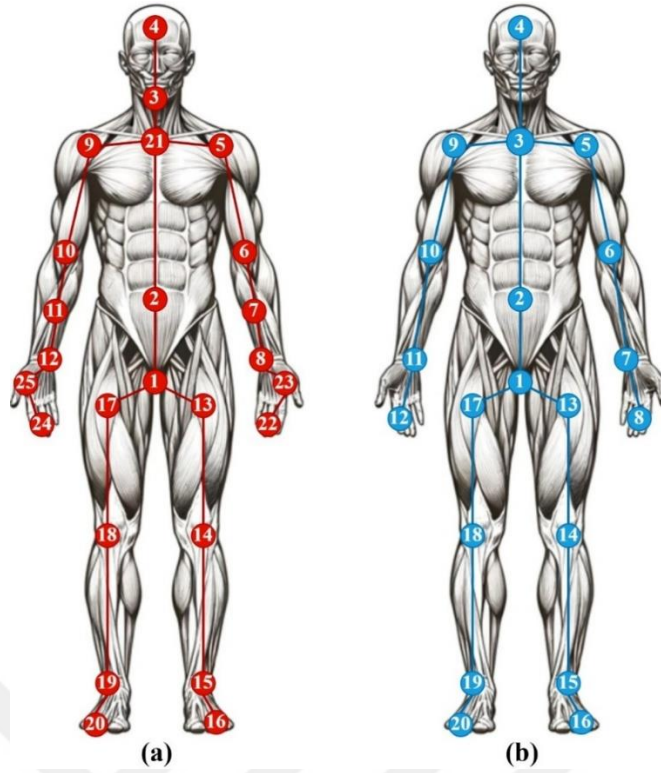
NTU-RGB+D 60 veri kümesi

NTU-RGB+D 60 (NTU-60) veri kümesi Shahroudy *et al.* (2016) tarafından 40 farklı denekten 60 insan eylemi kategorisinde toplanan 56.880 adet video klip içeren büyük ölçekli bir eylem tanıma veri kümesidir. Bu veri kümesinde her eylem örneği RGB, derinlik, kızılötesi ve 3B iskelet verileri olmak üzere dört farklı veri türüne sahiptir. Şekil 9'da NTU-RGB+D veri kümesinden örnek kareler verilmiştir. Şekilde ilk satır deneklerin sınıf içi performanslarını ikinci satır ise RGB, RGB+Eklemler, Derinlik, Derinlik+Eklemler ve kızılötesi veri türlerini göstermektedir.



Şekil 9. NTU-RGB+D veri kümesinden örnek kareler

Bu doktora çalışması kapsamında sadece 3B iskelet verileri kullanılmıştır. İskelet verileri, farklı ayarlara sahip üç adet Microsoft Kinect V2 kamerasından elde edilmiştir. Her eylem örneğinin maksimum çerçeve sayısı 300'dür ve her denek iskeleti için 25 adet vücut eklemi koordinatı yakalanmaktadır. NTU RGB+D veri kümesine ait 25 vücut eklemi topolojisi Şekil 10 (a)'da görselleştirilmiştir.



Şekil 10. (a) NTU RGB+D veri kümesine ait 25 vücut ekleminin topolojisi (b) Northwestern-UCLA veri kümesine ait 20 vücut ekleminin topolojisi.

NTU-RGB+D 120 veri kümesi

NTU-RGB+D 120 (NTU-120) (Liu *et al.* 2019) veri kümesi, NTU-60 veri kümesinin genişletilmiş versiyonudur. Daha fazla denek ve eylem kategorisi içerdiğinden çok daha zorlu bir kümesidir. Geniş bir yaş dağılımı aralığındaki insanlardan alınan 114.480 adet video klip içermektedir. Farklı kamera yüksekliği ve mesafesine sahip 32 farklı kamera kurulumu vardır ve her kurulum için bir kurulum kimliği verilmiştir. Veri kümesinde denek sayısı 106'ya, eylem kategorisi sayısı 120'e çıkmış ancak iskelet topolojisi aynı kalmıştır. NTU-60 ve NTU-120 veri kümelerinde yer alan eylem sınıfları günlük eylemler, karşılıklı eylemler ve tıbbi durumlar olmak üzere üç ana kategoriye ayrılmaktadır. Tablo 2'de E1 ile E60 arası etiketlenen eylemler NTU-60 veri kümesini, E1 ile E120 arasında etiketlenen eylemler ise NTU-120 veri kümesini oluşturmaktadır.

Tablo 2. NTU-60 ve NTU-120 Veri Kümesinde Yer Alan Eylem Sınıfları ve Kategorileri

Günlük Eylemler	
E1: Su içmek	E2: Yemek/atıştırmalık yemek
E3: Diş fırçalamak	E4: Saç taramak
E5: Bir şeyi düşürmek	E6: Bir şeyi yerden almak
E7: Bir şeyi fırlatmak	E8: Oturmak
E9: Oturur pozisyonundan kalkmak	E10: Alkışlamak

Tablo 2. (devamı)

E11: Kitap/gazete okumak	E12: Yazı yazmak
E13: Kâğıt yırtmak	E14: Ceket giymek
E15: Ceket çıkarmak	E16: Ayakkabı giymek
E17: Ayakkabı çıkarmak	E18: Gözlük takmak
E19: Gözlük çıkarmak	E20: Şapka/kep takmak
E21: Şapka/kep çıkarmak	E22: Neşelenmek
E23: El sallamak	E24: Bir şeye tekme atmak
E25: Cebe bir şey koymak / çıkarmak	E26: Tek ayak üzerinde zıplamak
E27: Yukarı zıplamak	E28: Telefonla konuşmak/cevap vermek
E29: Telefon/tabletle oynamak	E30: Klavyede yazı yazmak
E31: Bir şeyi parmakla göstermek	E32: Selfie çekmek
E33: Saate bakmak	E34: İki eli ovuşturmak
E35: Baş sallamak/baş eğmek/Evet demek	E36: Başını sallamak (hayır demek)
E37: Yüzü silmek	E38: Selam durmak
E39: Avuç içlerini birleştirmek	E40: Elleri önde çaprazlamak (dur demek)
E61: Kulaklık takmak	E62: Kulaklık çıkarmak
E63: Potaya şut atmak	E64: Top sektirmek
E65: Tenis raketiyle vurmak	E66: Pinpon topu sektirmek
E67: Sessiz ol işareti yapmak	E68: Saçı savurmak
E69: Baş parmak yukarı (beğeni)	E70: Baş parmak aşağı (beğenmeme)
E71: Tamam işareti yapmak	E72: Zafer işareti yapmak
E73: Kitap zımbalamak	E74: Para saymak
E75: Tırnak kesmek	E76: Makasla kâğıt kesmek
E77: Parmak şıklatmak	E78: Şişe açmak
E79: Koklamak	E80: Çömelmek
E81: Yazı tura atmak	E82: Kâğıt katlamak
E83: Kâğıdı buruşturmak	E84: Rubik küp oynamak
E85: Yüze krem sürmek	E86: Elin arkasına krem sürmek
E87: Çanta takmak	E88: Çanta çıkarmak
E89: Çantaya bir şey koymak	E90: Çantadan bir şey çıkarmak
E91: Kutu açmak	E92: Ağır nesneleri taşımak
E93: Yumruk sallamak	E94: Şapka/kep havaya fırlatmak
E95: Elleri havaya kaldırmak (iki el)	E96: Kolları çapraz yapmak
E97: Kollarla daire çizmek	E98: Kolları sallamak
E99: Yerinde koşmak	E100: Kalçaya tekme (geri tekme)
E101: Çapraz ayak parmağına dokunmak	E102: Yandan tekme atmak

Tablo 2. (devamı)

Tıbbi Durumlar	
E41: Hapşırma/öksürmek	E42: Sendelemek
E43: Düşmek	E44: Başa dokunmak (baş ağrısı)
E45: Göğse dokunmak (mide/kalp ağrısı)	E46: Sırta dokunmak (bel ağrısı)
E47: Boyuna dokunmak (boyun ağrısı)	E48: Mide bulantısı veya kusma
E49: El ya da kağıtla yelpazelenmek	E103: Esnemek
E104: Gerinmek	E105: Burnu silmek
Karşılıklı Eylemler / İki Kişilik Etkileşimler	
E50: Başka bir kişiye yumruk/şaplak atmak	E51: Başka bir kişiye tekme atmak
E52: Başka bir kişiyi itmek	E53: Başka bir kişinin sırtını sıvazlamak
E54: Başka bir kişiyi parmakla işaret etmek	E55: Başka birine sarılmak
E56: Başka birine bir şey vermek	E57: Başka birinin cebine dokunmak
E58: El sıkışmak	E59: Birbirine doğru yürümek
E60: Birbirinden uzaklaşarak yürümek	E106: Bir nesneyle başka bir kişiye vurmak
E107: Başka bir kişiye bıçak sallamak	E108: Vücutla çarpıp kişiyi devirmek
E109: Başka birinin eşyasını kapmak	E110: Başka birine silahla ateş etmek
E111: Ayağına basmak	E112: Beşlik çakmak
E113: Şerefe yapıp içmek	E114: Başka bir kişiyle bir şey taşımak
E115: Başka birinin fotoğrafını çekmek	E116: Başka birini takip etmek
E117: Başka birinin kulağına fısıldamak	E118: Başka biriyle bir şey takas etmek
E119: Birine destek olmak (elle tutarak)	E120: Taş-kâğıt-makas oyunu oynamak

Northwestern-UCLA veri kümesi

Northwestern-UCLA (N-UCLA) (Wang *et al.* 2014) veri kümesi, üç Kinect kamerası tarafından aynı anda birden fazla bakış açısından yakalanmış 10 eylem kategorisini (Tek elle bir şey kaldırmak, İki elle bir şey kaldırmak, Çöp atmak, Etrafta yürümek, Oturmak, Ayağa kalkmak, Giyinmek, Çıkarmak, Fırlatmak, Taşımak) kapsayan 1494 video klip içermektedir. Her eylem 10 farklı denek tarafından gerçekleştirilmiştir. N-UCLA veri kümesine ait 20 vücut ekleminin topolojisi Şekil 10 (b)'de görselleştirilmiştir. Tablo 3'de kullanılan veri kümelerinin örnek sayısı, sınıf sayısı, denek sayısı, sensör türü ve sağladıkları veri türleri verilmiştir.

Tablo 3. Kullanılan Veri Kümelerinin Özellikleri

Veri Kümeleri	Örnek Say.	Sınıf Say.	Denek Say.	Sensör	Veri Türleri
N-UCLA (Wang <i>et al.</i> 2014)	1.474	10	10	Kinect v1	RGB+D+3DJOINTS
NTU-60 (Shahroudy <i>et al.</i> 2016)	56.880	60	40	Kinect v2	RGB+D+IR+3DJOINTS
NTU-120 (Liu <i>et al.</i> 2019)	114.480	120	106	Kinect v2	RGB+D+IR+3DJOINTS

Veri Ön işleme

Önişlem adımları bireyin vücut yapısı, konumu ve kamera bakış açısından kaynaklanan farklılıkları azaltarak iskelet verilerinin standart bir temsilini sağlamaktadır. Böylece tutarlı ve doğru özelliklerin çıkarılmasını kolaylaştırarak eylem tanıma sistemlerinin güvenilirliğini ve hassasiyetini artırmaktadır.

Ötelemekten bağımsızlık

Kamera, hareketleri çeşitli açılardan ve konumlardan kaydedebilir ya da bir denek aynı eylemi farklı başlangıç noktalarından gerçekleştirebilir. Farklı kamera konumlarının getirdiği bu çeşitliliği minimize etmek için, bir çözüm yolu iskeletin koordinat sistemini, kamera merkezinden belirlenmiş bir eklem noktasına doğru ötelemektir (Shahroudy *et al.* 2016; Agahian *et al.* 2020). Bu noktada öteleme eklemi olarak merkez omurga eklemi (m_omurga) (Şekil 10 iki numaralı eklem) seçmek mantıklıdır. Çünkü omurga merkezi vücudun üst ve alt yarısını birbirine bağlar ve bir kişinin genel pozisyonunu, yönelimini ve hareketini yansıtan önemli bir eklemdir. Öteleme eklemi olarak omurga eklemi seçimi, bedenin genel konumunu ve yönelimini standart hale getirerek, dönüşüm işlemlerini daha tutarlı ve anlamlı hale getirecektir. Farklı kamera açılarından kaydedilen iskelet verilerindeki hareket analizi sonuçlarını iyileştirerek hareketi daha genel bir referans çerçevesinde değerlendirmeye olanak sağlayacaktır. Bu nedenle, ön işleme aşamasında ilk adım olarak, her bir çerçeve için koordinat sisteminin merkezini kamera koordinatlarından kişinin omurga eklemine taşınır. Bu dönüşüm, iskelet eklem pozisyonlarının kişinin konumuna bağlı olmamasını sağlar. Ötelemekten bağımsızlık için uygulanan dönüşüm Denklem (1)'de verilmiştir.

$$(x'_i, y'_i, z'_i) = (x_i - x_{m_omurga}, y_i - y_{m_omurga}, z_i - z_{m_omurga}) \quad (1)$$

Denklemden i eklem indeksini göstermektedir. Bu dönüşüm ile iskelet eklem konumları kişinin pozisyonundan bağımsız hale getirilmesi sağlanmaktadır.

Kamera bakış açısından bağımsızlık

Eylemler, kamera konumlarına bağlı olarak farklı bakış açılarından farklı gözükabilir. Bu nedenle, iskelet eklemlerinin kamera bakış açılarından bağımsız hale getirilmesi gerekmektedir. Bu işlem için eksenlerin 3B dönüşümler ile düzenlenmesi gerekmektedir. X eksenini, sağ omuzdan sol omuza uzanan 3B vektöre paralel olarak hizalanırken, Y eksenini, omurga tabanından omurgaya uzanan 3B vektör doğrultusunda yönlendirilir. Yeni X ve Y eksenlerinin çapraz çarpımı ($X \times Y$), Z eksenini olarak sabitlenir (Shahroudy *et al.* 2016). Bu

işlem, iskeletin üst kısmının yatay olarak hizalanmasını ve omurga boyunca düzgün bir şekilde yükselmesini sağlayarak görüntüleme açılarından kaynaklanan değişkenliği azaltır.

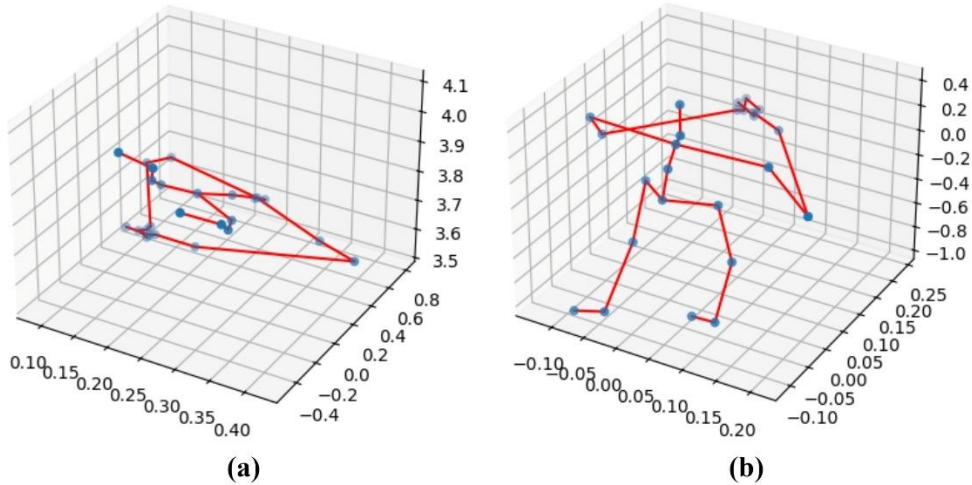
Ölçekten bağımsızlık

Genellikle eylemleri gerçekleştiren bireyler, farklı vücut ölçülerine sahiptir. Güçlü eylem modelleri oluşturabilmek için, farklı vücut ölçülerine sahip bireylerin eylem öznelikleri arasındaki tutarlılık sürdürülmelidir. Örneğin, iskeletten elde edilen bir değeri kullanarak, sabit bir aralığa ölçeklendirme işlemi gerçekleştirilebilir. Tüm eklem noktaları, Denklem (2), (3) omurga ve omurga tabanı eklemleri arasındaki mesafe temel alınarak ölçeklendirilir. Denklem (2)'de d omurga ile omurga tabanı eklemleri arasındaki mesafeyi ifade etmektedir. Denklem (3) aracılığıyla tüm eklem noktaları bu mesafeye göre ölçeklendirilmektedir.

$$d = \sqrt{(x_o - x_{o_tabanı})^2 + (y_o - y_{o_tabanı})^2 + (z_o - z_{o_tabanı})^2} \quad (2)$$

$$(x'_i, y'_i, z'_i) = \left(\frac{x_i}{d}, \frac{y_i}{d}, \frac{z_i}{d} \right) \quad (3)$$

Şekil 11'de veri ön işlemenin bir örneğini verilmiştir. Şekil 11 (a)'da su içme eylemini temsil eden ham eklem verileri gösterilmiştir. Bu veriler, öteleme, kamera bakış açısı ve ölçek gibi etmenlerden arındırılarak normalize edilmiş ve elde edilen sonuçlar Şekil 11 (b)'de sunulmuştur. Bu örnek, ön işlemenin sağladığı iyileştirmeyi açık bir şekilde ortaya koymaktadır.



Şekil 11. NTU-RGB+D veri kümesindeki "su içme" eylemi için bir veri ön işleme örneği. (a) Orijinal iskelet verisi (b) Ön işleme adımlarından sonra iskelet verisi

Çoklu Modalite Temsili

Eklem modalitesi

İskelet dizileri vücut eklemlerinin 3B koordinat bilgilerinin bir koleksiyonu ile temsil edilir. T çerçeveli bir iskelet dizisinde, t . çerçevedeki i . vücut eklemine koordinatları Denklem (4) ile ifade edilir.

$$E_t^i = (x_t^i, y_t^i, z_t^i) \quad t \in [1, \dots, T], i \in [1, \dots, N] \quad (4)$$

T toplam çerçeve sayısını N ise bir çerçevedeki vücut eklemlerinin sayısını göstermektedir. t . çerçevedeki iskelet verileri $S_t = \{E_t^1, E_t^2, \dots, E_t^N\}$ olarak gösterilir. Tüm iskelet koleksiyonu Denklem (5) ile ifade edilmektedir.

$$C = \{S_1, S_2, \dots, S_T\} \quad (5)$$

Kemik modalitesi

Literatürde bulunan mevcut çalışmalar iskelet tabanlı eylem tanımda eklem bilgileri ile kemik bilgilerinin bir araya getirilmesinin önemini göstermiştir (Shi *et al.* 2019a, 2019b). Her kemik iki eklemle bağlı olduğundan, iskeletin ağırlık merkezine yakın olan eklem kaynak eklem, ağırlık merkezinden uzak olan eklem ise hedef eklem olarak tanımlanır. Her kemik, yalnızca uzunluk bilgisini değil aynı zamanda yön bilgisini de içeren, kaynak ekleminden E_i hedef eklemine E_j işaret eden bir vektör olarak temsil edilir. t . çerçevede kaynak eklem E_i^t hedef eklem ise E_j^t olduğundan kemik bilgisi olan K_{ij}^t Denklem (6) ile hesaplanmaktadır.

$$K_{ij}^t = (x_i^t - x_j^t, y_i^t - y_j^t, z_i^t - z_j^t), \quad \forall i \in N, t \in T \quad (6)$$

İskelet çizgesinde döngü bulunmadığı için her bir kemiğe benzersiz bir hedef eklem atanabilmektedir. Merkez eklem herhangi bir kemiğe atanmadığı için eklem sayısı kemik sayısından bir fazladır. Eklem ve kemik modalitelerinde standartlaşma sağlamak için merkezi ekleme sıfır değeri olan boş bir kemik eklenmektedir.

Hareket modalitesi

Eklem hareketi (EH) ve kemik hareket (KH) bilgileri ardışık çerçeveler arasındaki eklem ve kemik pozisyonlarındaki değişiklik olarak hesaplanmaktadır. Yani, bir çerçevenin eklem koordinatlarından bir sonraki çerçevenin eklem koordinatları çıkarılarak eklem hareket bilgileri Denklem (7)'de verildiği gibi elde edilir.

$$EH_i^t = E_i^{t+1} - E_i^t, \quad \forall i \in N, t \in T \quad (7)$$

Aynı şekilde bir çerçevenin kemik koordinatlarından bir sonraki çerçevenin kemik koordinatları çıkarılarak kemik hareket bilgileri Denklem (8)'de verildiği gibi elde edilir.

$$KH_i^t = K_i^{t+1} - K_i^t, \quad \forall i \in N, t \in T \quad (8)$$

Son çerçevenin bir sonraki çerçevesi olmadığından dolayı eklem ve kemik hareket bilgisi 0 olarak ayarlanır. Hareket bilgilerinin elde edilmesindeki amaç bu bilgileri kullanarak geliştirilen ağ yapısı ile eylemlere ait zamansal hareket bağımlılıkları hakkındaki özellikleri çıkarmaktır.

Uzaysal-Zamansal Çizge Yapısı

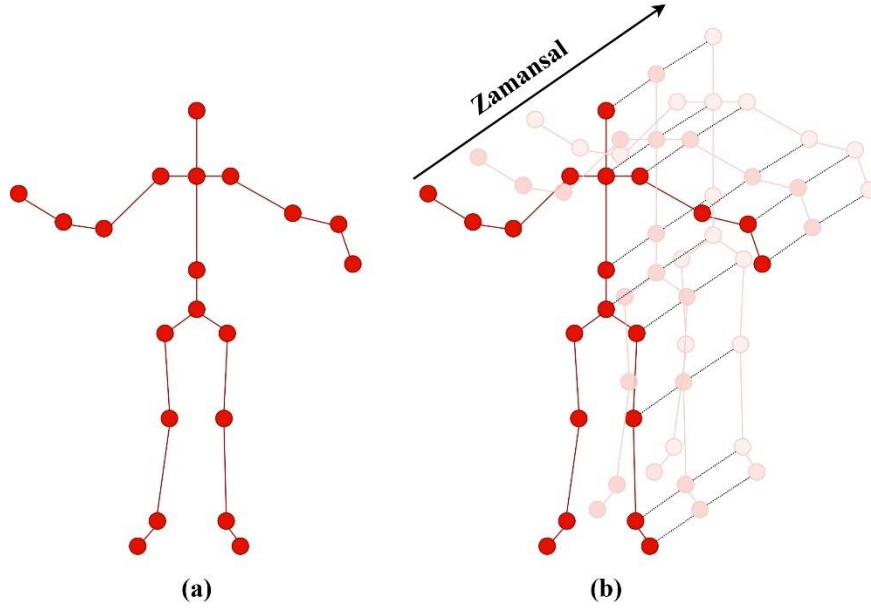
İnsan iskeleti, eklem noktaları ve bu noktalar arasındaki bağlantılarla temsil edilen bir yapıdır. Bu yapı, insan hareketlerinin modellenmesi ve analizi için temel bir çerçeve sağlar. İskelet tabanlı çizge modellerinde eklem noktaları düğümler kümesini (V), bu noktalar arasındaki bağlantılar ise kenarlar kümesini (E) oluşturur. Düğüm noktalarında, eklemlerin 2B/3B koordinat bilgileri bulunmaktadır. Hem vücut içi hem de çerçeveler arası bağlantıya sahip N eklemlili ve T çerçeveli bir iskelet dizisi üzerinde uzaysal-zamansal çizge $G = (V, E)$ şeklinde tanımlanır. Çizge yapısındaki düğümler kümesi V , Denklem (9)'da verildiği gibi tüm eklem noktalarını içermektedir.

$$V = \{v_{ti} \mid t = 1, \dots, T, \quad i = 1, \dots, N\} \quad (9)$$

Denklem 9'daki N , toplam eklem sayısını ifade etmektedir. İskelet dizileri üzerinde uzaysal-zamansal çizge iki adımda oluşturulur. İlk olarak, bir çerçeve içindeki eklemler insan vücut yapısının fiziksel bağlanabilirliğine göre kenarlarla birleştirilir. Daha sonra her eklem ardışık çerçevede aynı eklem yerine bağlanır. Uzaysal-zamansal çizge yapısında kenar kümesi E , uzaysal kenarlar ve zamansal kenarlar olmak üzere iki alt kümeden oluşmaktadır (Yan *et al.* 2018).

Uzaysal kenarlar

Tek bir zaman dilimindeki düğümler arasındaki ilişkileri ifade eder ve $E_s = \{v_{ti}v_{tj} \mid (i, j) \in H\}$ şeklinde tanımlanır. H , doğal olarak bağlı insan vücudu eklemlerinin kümesini temsil etmektedir. Uzaysal kenarlar, insan vücut yapısında eklemlerin fiziksel yakınlıklarına dayanarak belirlenmektedir. Uzaysal kenarlar Şekil 12 (a)'da görselleştirilmiştir.



Şekil 12. İskelet dizisinin uzaysal-zamansal çizge gösterimi: (a) uzaysal çizge, (b) uzaysal-zamansal çizge

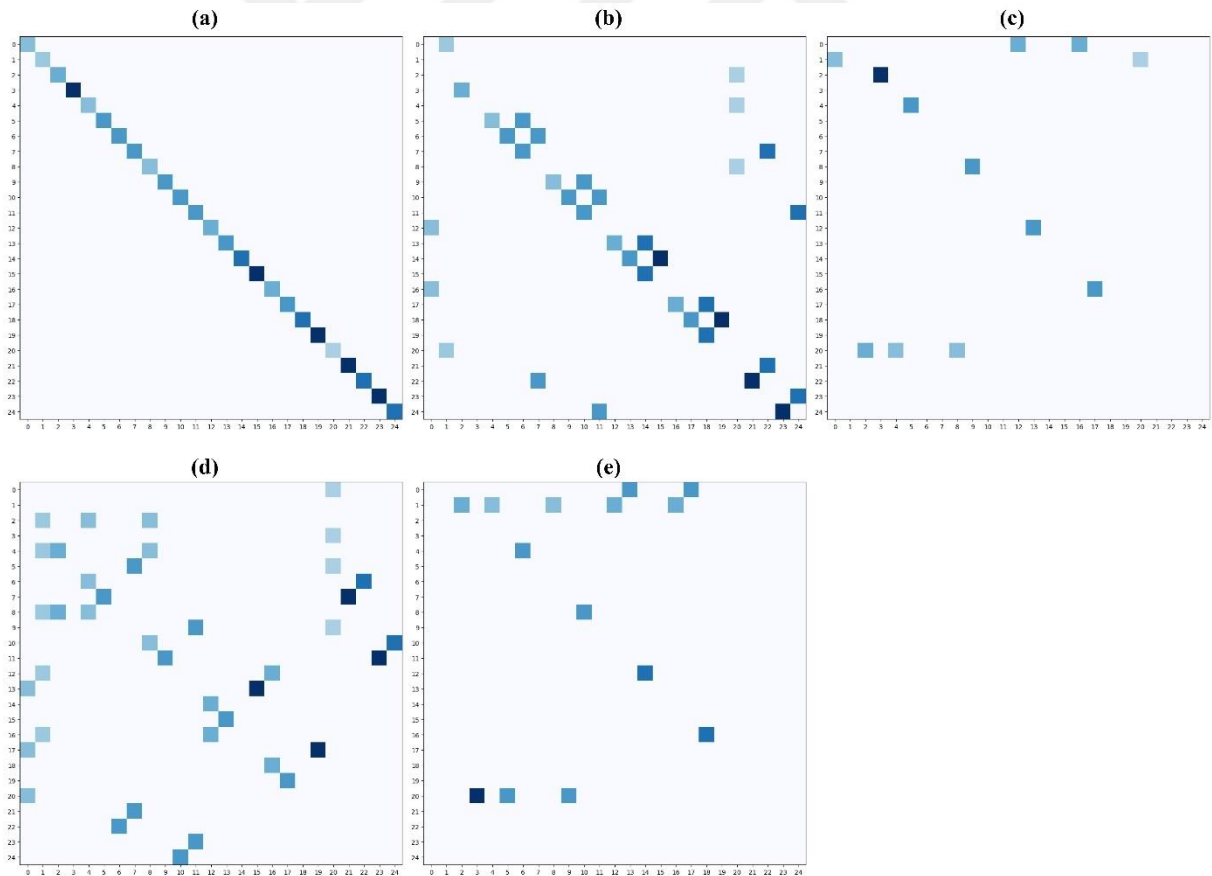
Zamansal kenarlar

Düğümlerin zaman içerisindeki evrimini ifade eder ve $E_F = \{v_{ti}v_{(t+1)i}\}$ şeklinde tanımlanır. Zamansal kenarlar, aynı eklemleri ardışık zaman dilimlerinde birbirine bağlar. Bu kenarlar, bir eklemin durumunun zaman içindeki değişimini gösterir ve hareketin zamansal bağımlılıklarını ve korelasyonlarını kodlamaya olanak tanır. Zamansal kenarlar Şekil 12 (b)'de görselleştirilmiştir. Sonuç olarak, uzaysal kenarlar aynı zaman diliminde eklem noktaları arasındaki ilişkileri tanımlarken, zamansal kenarlar bir eklemin ardışık zaman dilimlerindeki hareketini modellemeye olanak tanır. Bu şekilde oluşturulan uzaysal-zamansal çizge yapısı hem eklemler arasındaki yerel bağlantıları hem de zaman içerisindeki dinamik değişimleri kapsayarak, hareketlerin daha etkili bir şekilde temsil edilmesini sağlar.

Bitişiklik matrisleri

ST-GCN modelleri, uzaysal-zamansal çizgeler üzerinde evrişim işlemlerini yapmak için bitişiklik matrislerini kullanmaktadır. Bitişiklik matrisleri, her bir eklemin diğer eklemlerle olan ilişkisini ve bu ilişkilerin bir kenar olup olmadığını ifade eder. ST-GCN, uzaysal ilişkileri modellemek bitişiklik matrislerini kullanarak çizge üzerinde evrişim işlemlerini gerçekleştirirken, zamansal ilişkileri modellemek için de zaman boyunca düğümlerin durumlarını takip eder. Yani, bitişiklik matrisi yalnızca uzaysal ilişkileri temsil eder, ancak ST-GCN bu matrisi zamansal boyutunu da içerecek şekilde genişletir. Böylece model, eklemler arasındaki ilişkileri ve bu ilişkilerin zamanla nasıl değiştiğini öğrenir.

Uzaysal bilgileri işlemek için uzaysal evrişim uygulamaları, bu evrişim hangi komşuluk seviyesinde yapılacağını belirleyen parametre ise atlama boyutudur (hop size). Atlama boyutu bir eklemenin diğer eklemelerle bağlantı kurma derecesini ifade eder ve hop terimi ile gösterilmektedir. Örneğin, model 1-hop'da yalnızca doğrudan bağlantılı komşuları dikkate alırken, 2-hop'da ise ikinci dereceden komşuları da hesaba katmaktadır. Atlama boyutu değeri arttıkça, modelin daha geniş bir bağlamı öğrenmesi sağlanır. Ancak, bu aynı zamanda hesaplama maliyetini ve modelin karmaşıklığının da artmasına yol açmaktadır. Bu doktora çalışmasında geliştirilen mimarilerde G_{25} çizge yapısı için atlama boyutu değeri 2 olarak, G_{11} ve G_6 çizge yapıları için ise 1 olarak belirlenmiştir. Bu seçim, her bir çizge yapısının yapısal özelliklerine ve komşuluk ilişkilerinin derinliğine göre yapılmıştır. G_{25} yapısındaki daha fazla komşuluk ilişkisinin, daha geniş bir bağlamın öğrenilmesine ihtiyaç duyduğu için hop size değeri 2 olarak belirlenirken, G_{11} ve G_6 yapılarında daha sınırlı bir komşuluk ilişkisi ile daha lokal öğrenme yeterli görülmüştür. Şekil 13'de G_{25} çizge yapısı için geliştirdiğimiz mimariler tarafından kullanılan normalleştirilmiş sabit bitişiklik matrisleri görselleştirilmiştir. Bitişiklik matrislerinin normalleştirilmesi işlemi ilerleyen bölümlerde anlatılmıştır.



Şekil 13. Sabit olarak kullanılan bitişiklik matrisleri: (a) 0-hop bağlantılar, (b) ve (c) 1-hop atlama mesafesinde olan ve sırasıyla merkez düğüme daha yakın ve daha uzak olan düğümler arasındaki bağlantılar, (d) ve (e) sırasıyla merkez düğüme daha yakın ve daha uzak olan 2-hop atlama mesafesinde olan düğümler arasındaki bağlantıları temsil etmektedir

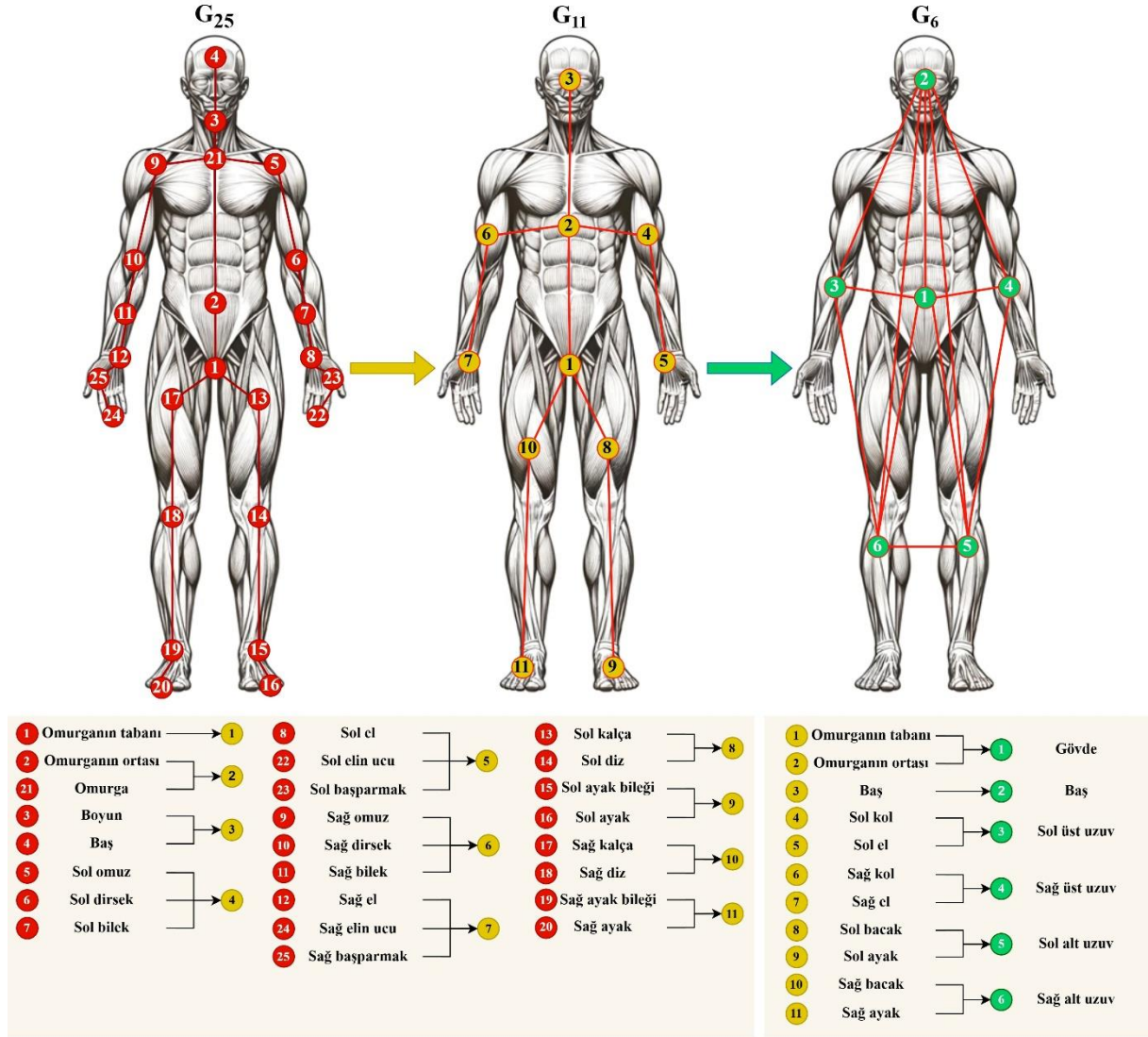
Bitişiklik matrisleri hem satırların hem de sütunların iskelet eklemlerini temsil ettiği kare matrislerdir. Matrislerdeki (i, j) konumundaki her bir eleman, i eklemi ile j eklemi arasındaki bağlantının ağırlığını temsil eder. Matristeki her bir elemanın mavi ölçeği, bu ağırlığın gücünü göstermektedir. Atlama boyutu değeri 2'dir. Bitişiklik matrisleri 0, 1 ve 2 atlama boyutlarıyla fiziksel olarak bağlı olan eklemler arasındaki bağımlılıkları temsil eden seyrek matrislerdir. Bu sabit bitişiklik matrisleri, Shiraki *et al.* (2020)'deki uzaysal stratejisi kullanılarak oluşturulmuştur. Her matris, düğümler arasındaki mesafeye ve merkez düğüme olan mesafeye göre farklı kategorilere ayrılmıştır. (a) matris, 0-hop atlama mesafesinde olan bağlantıları temsil eder. (b) ve (c) matrisleri, 1-hop atlama mesafesinde olan ve sırasıyla merkez düğüme daha yakın ve daha uzak olan düğümler arasındaki bağlantıları temsil eder. (d) ve (e) matrisleri, sırasıyla merkez düğüme daha yakın ve daha uzak olan 2-hop atlama mesafesinde olan düğümler arasındaki bağlantıları temsil eder. Bu şekilde, eklemler arasındaki ilişkiler daha ayrıntılı olarak modellenabilmektedir.

Çok Ölçekli Çizge Yapıları

İnsan eylemleri, kinematik olarak vücudun farklı bölgelerindeki eklem gruplarının uyumlu bir şekilde koordinasyonu sonucunda oluşur. Bu koordinasyon, motor becerilerimizin temelini oluşturur ve karmaşık hareketlerin doğru ve etkili bir şekilde gerçekleştirilmesini sağlar. Ayrıca, belirli eylemleri doğru bir şekilde tanımlamak ve sınıflandırmak için gerekli temel hareket kalıplarını da oluşturur. Bu karmaşık koordinasyon, eylemlerin doğru bir şekilde temsil edilmesi ve tanımlanması için kritik bir öneme sahiptir. Ancak, iskelet tabanlı eylem tanıma için literatürdeki mevcut yöntemlerin çoğu bu koordinasyonu göz ardı eder ve tek ölçekli çizge yapıları kullanır. Ancak tek ölçekli çizge yapılarını kullanan yöntemler, yalnızca sınırlı bir bağlamda yerel özellikleri çıkarmakla yetinirken, farklı vücut bölgeleri arasındaki daha geniş kapsamlı ve yüksek düzeyli anlamsal ilişkileri göz ardı etmektedir. Oysa bir eylemin başarılı bir şekilde tanınabilmesi, eklemler arası hem yerel hem de küresel ilişkilerin modellenmesini gerektirir. Bu eksiklik, özellikle karmaşık hareketlerin tanımlanmasında önemli bir performans kaybına yol açar. Örneğin, bir spor hareketi sırasında üst vücut, alt vücut ve gövde arasındaki koordinasyon hayati bir rol oynar. Bu bağlamda çok ölçekli yapılar, eklemlerin farklı gruplar arasında nasıl etkileşimde bulunduğunu anlamaya olanak tanır ve bu sayede eylemlerin daha hassas ve etkili bir şekilde tanımlanmasını sağlar.

Literatürde bazı çalışmalar (Yang *et al.* 2021; Feng *et al.* 2021) insan vücudundaki eklemler arasındaki ilişkileri daha iyi temsil edebilmek için çok ölçekli çizge yapıları kullanmayı önermiştir. Feng ve diğerleri, vücut-parça ilişkisi modelini kullanarak eklemleri komşuluk ilişkilerine göre gruplandırmış ve bu sayede çok ölçekli bir çizge yapısı

oluşturmuşlardır. Bu doktora çalışmasında geliştirilen mimarilerde, çok ölçekli çizge yapıları oluşturmak için bu yaklaşım benimsenmiştir. Vücut-parça ilişkisi modeli, iskelet yapısının doğal hiyerarşisini ve uzaysal bağıntılarını dikkate alarak, hareketleri daha ayrıntılı bir şekilde analiz etme olanağı sunmaktadır. Şekil 14’de vücut-parça ilişkisi modeline göre oluşturulmuş farklı ölçekteki çizge yapıları görselleştirilmiştir.

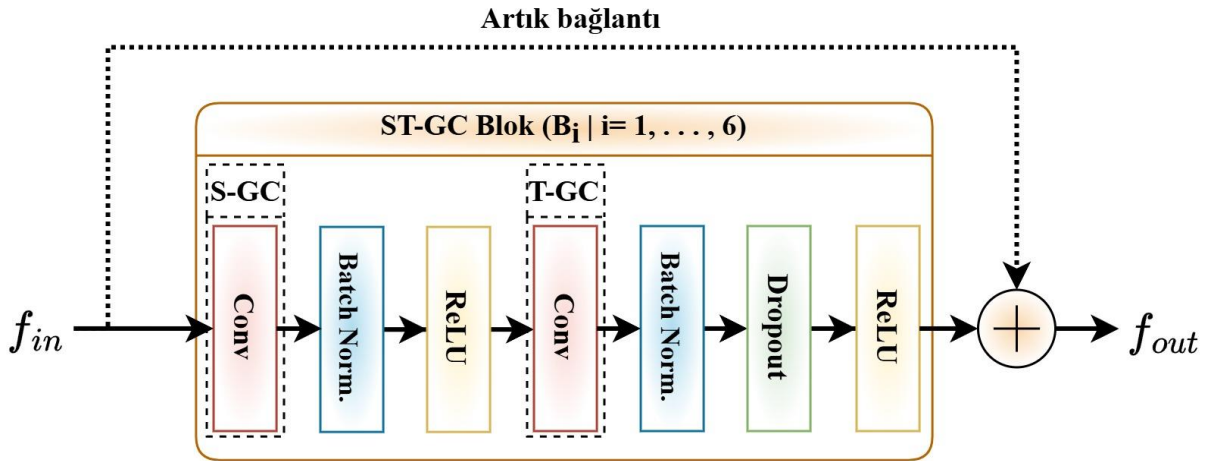


Şekil 14. Vücut-parça ilişkisi modeline dayalı olarak oluşturulan farklı ölçeklerdeki çizge yapıları (G₂₅, G₁₁ ve G₆)

Şekil 14’de görüldüğü gibi, insan vücudunu daha ayrıntılı ve anlamlı bir şekilde modellemek için G₂₅, G₁₁ ve G₆ olmak üzere üç çizge yapısı oluşturulmuştur. G₂₅ çizgesi, 25 eklemliyken, G₁₁ çizgesi vücudu 11 ana bölgeye, G₆ çizgesi ise vücudu 6 genel bölgeye bölmektedir. Bu yaklaşımlar, insan hareketlerinin daha doğru ve etkili bir şekilde tanımlanmasını ve sınıflandırılmasını sağlar. Çok ölçekli çizge yapıları ile eğitilen modeller, eylem tanıma doğruluğunu ve genelleme yeteneğini geliştirerek daha güvenilir sonuçlara ulaşmaktadırlar.

Uzaysal-Zamansal Çizge Evrişimsel Ağlar

ST-GCN ağ yapısı eylem tanıma, eylem tahmini ve davranış analizi gibi uygulamalarda insan iskelet verilerini modellemek için Yan *et al.* (2018) tarafından tasarlanmış bir çerçevedir. Bu ağlar, yığılmış uzaysal-zamansal blokların bir hiyerarşisi olarak yapılandırılmıştır ve insan iskeletinin doğal çizge yapısını kullanarak uzaysal-zamansal ilişkileri modellemektedir. ST-GCN, tasarlanan mimari yapıya göre birden fazla ST-GC bloğunun birleşiminden oluşmaktadır. Her bir blok, bir uzaysal çizge evrişimi (S-GC), iki toplu normalleştirme (Batch Normalization) katmanı, iki aktivasyon fonksiyonu (ReLU) ve bir zamansal evrişim (T-GC) katmanından meydana gelmektedir. ST-GC blokları, artık bağlantı yapısına sahiptir. Bu artık bağlantı sayesinde, her bir ST-GC bloğu, girdi verilerini işleyerek bir çıktı üretirken, girdi verileri doğrudan çıktıya eklenir. Bu yapı, modelin daha derin hale gelmesine olanak tanırken, eğitimin daha etkin bir şekilde gerçekleşmesini sağlamaktadır (Yan *et al.* 2018). ST-GC blok yapısı Şekil 15’de görselleştirilmiştir.



Şekil 15. ST-GC blok yapısı

Uzaysal çizge evrişimi

Uzaysal alt modül S-GC, her bir düğümün özelliklerini ona komşu düğümlerle birleştirerek hesaplar. Bu işlem, eklemler arasındaki yerel ilişkileri modellemek için, Kipf and Welling (2016) tarafından önerilen Denklem (10) ile gerçekleştirilir. Bu yöntem, düğümler arasında evrişimler gerçekleştirmek için çizge yapısını kullanır ve çizge içindeki uzaysal bağımlılıkları yakalar.

$$f_{out} = \sum_k^{K_s} (f_{in} A_k) W_k \quad (10)$$

Denklem (10) S-GC katmanının çıktı özelliği f_{out} 'un nasıl hesaplandığını göstermektedir. f_{in} giriş özellik matrisini, K_s bitişiklik matrislerinin sayısını, A_k k -inci

komşuluk için bitişiklik matrisinin normalleştirilmiş halini, W_k k -inci komşuluk için öğrenilebilir ağırlık matrisini göstermektedir. Bu denklem farklı bitişiklik matrisleri ve karşılık gelen ağırlık matrisi kullanılarak birden fazla çizge evrişim işleminin bir araya getirilip toplanması ile çıktının nasıl oluşturulduğunu göstermektedir. Denklem (11), normalleştirilmiş A_k bitişiklik matrisinin nasıl hesaplandığını göstermektedir.

$$A_k = D_k^{-\frac{1}{2}}(\tilde{A}_k + I)D_k^{-\frac{1}{2}}, \quad D_{ii} = \sum_k^{K_s} (\tilde{A}_k^{ij} + I_{ij}) \quad (11)$$

Denklem (9)'da $\tilde{A}_k$, k -inci komşuluk için bitişiklik matrisini yani çizgedeki bağlantıların temel yapısını temsil eder. I birim matrisi düğümlerin kendi kendilerine olan bağlantılarını eklemek için kullanılır ve her bir düğümün kendi bilgisini de öğrenmesini sağlamaktadır. D_k ise $\tilde{A}_k + I$ ile ilişkili derece matrisini temsil eder. Derece matrisi D_k , $\tilde{A}_k + I$ matrisindeki her satırdaki elemanların toplamını köşegen eleman olarak alan bir köşegen matristir. Bu şekilde, her düğümün derece bilgisi, yani o düğüme bağlı olan kenarların sayısı elde edilmektedir. Bu normalleştirme işlemi, her bir düğümün özelliklerinin komşu düğümlerden gelen bilgilerle ölçeklenmiş ve dengelenmiş bir şekilde birleştirilmesini sağlar. Bu özellikle derin çizge sınır ağlarında eğitimi stabilize etmek ve öğrenmeyi hızlandırmak için önemlidir. Normalleştirme işlemi, yönlendirilmemiş çizgeler için $D_k^{-\frac{1}{2}}(\tilde{A}_k + I)D_k^{-\frac{1}{2}}$ formülü ile gerçekleştirilir. Ancak, bu çalışmada yönlendirilmiş çizgeler kullanıldığı için normalleştirme faktörü, bir düğümden çıkan kenar sayısının tersi olarak belirlenmiştir. Yönlendirilmiş çizgeler için normalleştirme denklemi şu şekildedir:

$$A_k = D_k^{-1}(\tilde{A}_k + I) \quad (12)$$

Bu, bir düğümden diğerine geçiş olasılığını temsil eder. Bir düğümden çıkan daha fazla kenar varsa, her bir kenara geçiş olasılığı daha düşük olacaktır. Bu normalleştirme, özellik yayılımının düğüm dereceleri ile normalize edilmesini ve eğitim sürecinin kararlı hale gelmesine yardımcı olur. Şekil 16'da yönlendirilmiş çizgenin normalleştirme sözde kodu verilmiştir.

```

1: function Normalize_Digraph(A)
2:   DI'yi A matrisinin her sütununun toplamı olarak başlat
3:   A matrisinin yüksekliğini (h) ve genişliğini (w) al
4:   Dn'yi w x w boyutunda sıfır matrisi olarak başlat
5:   for i = 0 to w-1 do
6:     if DI[i] > 0: then
7:       Set Dn[i, i] to (DI[i])(-1)
8:     end if
9:   end for
10:  A'yı Dn ile çarparak AD'yi hesapla
11:  return AD
12: end function

```

Şekil 16. Yönlendirilmiş çizgenin normalleştirme işlemi sözde kodu

Zamansal evrişim

Zamansal evrişim T-GC'nin temel amacı, eklemlerin zaman içindeki dinamik değişimlerini ve sürekliliğini anlamaktır. Zamansal evrişimler, bir dizi çerçeveden gelen verileri birleştirerek zaman içindeki hareketleri öğrenmeye çalışır. Bu işlem 1-Boyutlu evrişim kullanılarak gerçekleştirilir. Zamansal evrişim, zaman boyutundaki hareketlerin dinamiklerini yakalamak için S-GC katmanının çıktı tensörü $F \in \mathbb{R}^{C \times T \times V}$, in $(T \times V)$ boyutları üzerinde $1 \times K_t$ boyutunda bir 1-Boyutlu evrişim olarak uygulanır. Burada C kanal sayısını, T çerçeve sayısını, V eklem sayısını ve K_t zaman boyutundaki çekirdek boyutunu temsil eder. T-GC, her evrişim işlemi sırasında K_t çerçeve boyunca özellik çıkarımı yapmaktadır. Sonuç olarak T-GC alt modülü, zaman boyutunda çerçeveler arasındaki ilişkileri analiz ederek hareketlerin zaman içindeki dinamiklerini çıkarır ve bu analizleri özellik vektörlerine dönüştürür.

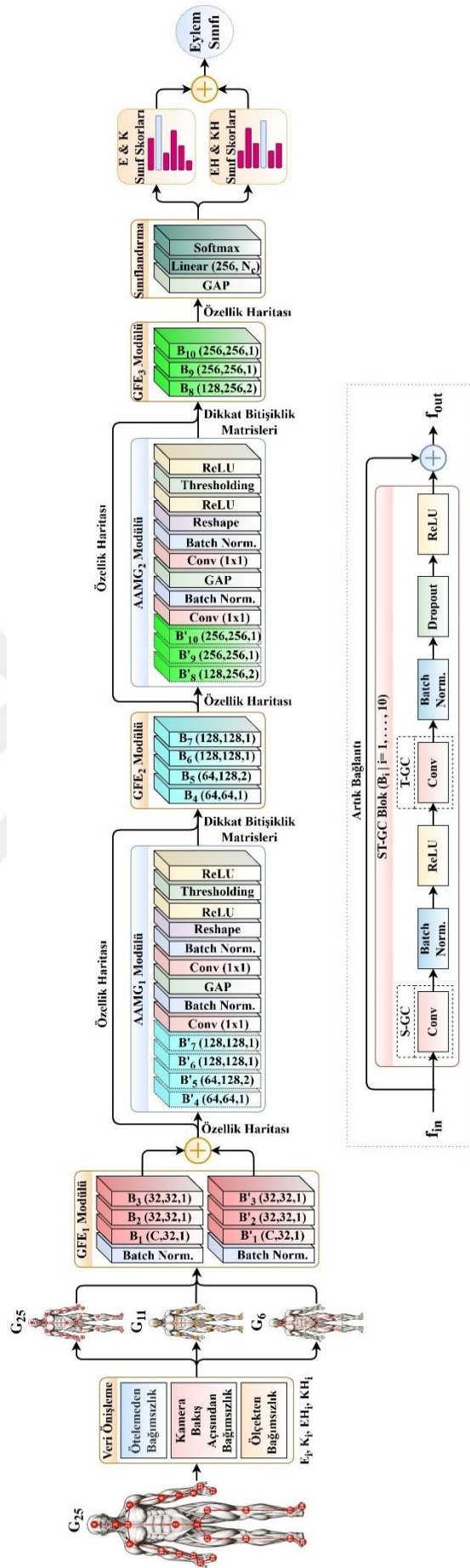
AGMS-GCN Mimarisi

Mimari yapıya genel bakış

İskelet tabanlı eylem tanıma için dikkat rehberliğinde çok ölçekli çizge evrişimli ağ yapısı AGMS-GCN'nin (Kilic *et al.* 2025) genel mimarisi Şekil 17'de verilmiştir. Önerilen mimari yapı incelediğinde iskelet verileri ilk olarak önışlem aşamasından geçmektedir. Eklem koordinatları koordinat sistemine ve ölçüm koşullarına bağlı olarak değişebilir. Örneğin, aynı hareketi gerçekleştiren iki kişi, hareketi farklı bir pozisyon veya açıda gerçekleştirdiğinde farklı koordinat değerlerine sahip olabilir. Bu nedenle, iskelet verilerini analiz etmek ve karşılaştırmak için verileri standart bir koordinat sistemine dönüştürmemiz gerekir. Bu işlem,

iskelet verilerinin veri ön işleme bölümünde belirtilen adımlara tabi tutulmasını içermektedir. Ön işleme adımı, farklı koordinat sistemlerinde veya farklı koşullar altında ölçülen iskelet verilerinin karşılaştırılabilir olmasını sağlar. Bu, iskelet verilerinin analizinde ve özellikle derin öğrenme modellerinin eğitiminde önemlidir çünkü derin öğrenme modelleri genellikle standartlaştırılmış giriş verileri gerektirir.





Şekil 17. AGMS-GCN'nin genel mimarisi B_i ile temsil edilen her ST-GC bloğunun iç yapısı, şeklin alt kısmında gösterilmiştir

İskelet verileri yalnızca iskelet eklemlerinin x, y, z koordinat bilgilerini içerir. Ön işleme adımından sonra eklem (E) koordinat modalitesi kullanılarak eklem hareketi (EH), kemik (K) ve kemik hareketi (KH) modaliteleri oluşturulur. Daha sonra bu modaliteler ve çizge yapısı birinci özellik çıkarma modülü olan GFE_1 (Graph Feature Extraction) girdi olarak verilir. GFE_1 modülünde bulunan ST-GC blokları, çizge yapısı ile iskelet verileri üzerinde uzaysal-zamansal çizge evrişim işlemleri gerçekleştirir ve düşük seviyeli daha genel özellikleri içeren özellik haritasını oluşturur. Kapsamlı ablasyon çalışmaları, eklem ve kemik modalitelerinin ilk özellik çıkarım aşamasından sonra özellik haritalarının birleştirilmesinin daha iyi sonuçlar verdiğini göstermiştir. Aynı durum eklem hareket ve kemik hareket modaliteleri için de geçerlidir. Bu nedenle GFE_1 modülünde iki akışlı bir yapı bulunmaktadır. İlk akışta, eklem modalitesi üzerine ST-GC blokları uygulanarak özellik haritaları çıkarılır. İkinci akışta ise kemik modalitesi üzerine ST-GC blokları uygulanarak özellik haritaları oluşturulur ve bu haritalar birleştirilir. Bu aşamadan sonra dikkat bitişiklik matrisi üretici modülü ($AAMG_1$) geçici ST-GC blokları uygulayarak daha yüksek seviyeli özellikler elde eder. Ardından, özellik haritasına dikkat mekanizması uygulanarak, her hareket için en uygun eklem ilişkilerini temsil eden dikkat bitişiklik matrisleri oluşturulur. Dikkat tabanlı bitişiklik matrislerini oluşturmak için dinamik eşikleme mekanizması kullanılır. Bu mekanizma, özellik haritasının ortalamasına dayalı bir eşik değeri hesaplar. Eşik değerinin altında kalan değerler sıfırlanarak daha az önemli özellikler baskılanırken, önemli özellikler vurgulanır. Bu adaptif eşikleme süreci, en önemli uzaysal ve zamansal ilişkileri yansıtan dikkat bitişiklik matrislerinin oluşturulmasını sağlamaktadır.

GFE_2 özellik çıkarım modülü, $AAMG_1$ modülünden gelen dikkat bitişiklik matrisleri, sabit bitişiklik matrisleri ve GFE_1 modülünden gelen özellik haritasını alarak ST-GC bloklarıyla uzaysal-zamansal çizge evrişim işlemleri gerçekleştirir. Böylece, daha üst düzey özellikler içeren yeni bir özellik haritasını oluşturulur. Bu özellik haritası kullanılarak $AAMG_2$ modülünde bitişiklik matrisleri güncellenir.

GFE_3 modülü, $AAMG_2$ modülünden gelen dikkat bitişiklik matrisleri ve GFE_2 modülünden gelen özellik haritalarını kullanarak daha yüksek seviyeli ve soyut özellikler içeren yeni özellik haritasını oluşturur. AGMS-GCN yapısında, özellik haritaları kademeli olarak dikkat bitişiklik matrisleriyle geliştirilir.

Son olarak, özellik haritasına sınıflandırma modülünde küresel ortalama havuzlama (Global Average Pooling (GAP)) katmanı uygulanarak 256 boyutlu bir özellik vektörü elde edilir. Daha sonra, tam bağlı bir doğrusal katman ve Softmax fonksiyonu uygulanarak sınıf skorları hesaplanır. Eklem hareket ve kemik hareket modaliteleri içinde aynı işlemler gerçekleştirilir. E & K ve EH & KH sınıf skorları toplanır. Bu yapı, her ölçek seviyesi için ayrı

ayrı çalışarak nihai sınıf bilgisi için ölçek sınıf skorları toplanır. AGMS-GCN mimarisi, 3B iskelet verisinin temelinde bulunan gizli bilgileri etkili bir şekilde kodlar, özellikle eklem hareketi örüntüleri ve bunların ilişkilerini aşamalı olarak işleyerek güçlü uzaysal-zamansal bilgileri çıkarmayı sağlar. Tüm AGMS-GCN mimarisi python programlama dili kullanılarak, grafik işlem birimi ile güçlü ivmeli tensör hesaplama yapabilen PyTorch kütüphanesi ile inşa edilmiştir.

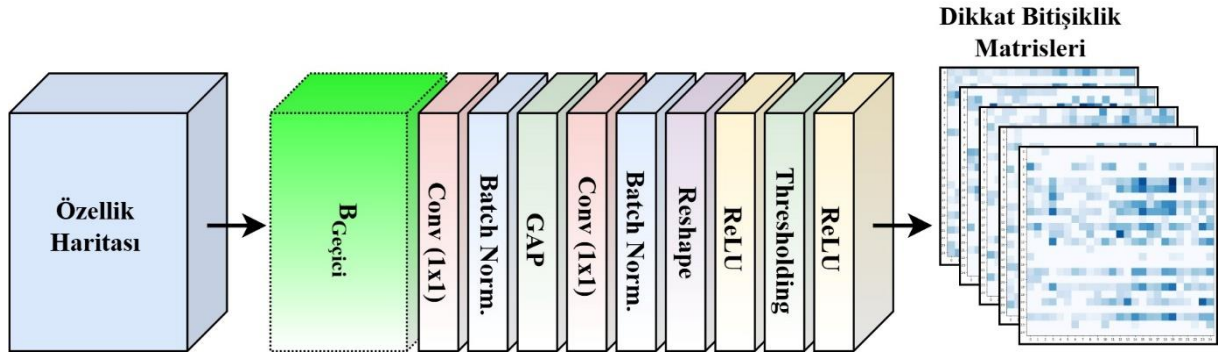
Çizge özellik çıkarıcı modülü

AGMS-GCN mimarisi, her biri ST-GC blokları içeren üç özellik çıkarıcı modülünden ($GFE_{1,2,3}$) oluşmaktadır. Bu ST-GC blokları, insan eylemlerini tanımak için zaman içindeki iskelet verilerinin uzaysal-zamansal özelliklerini bir arada modellemektir. İskelet eklemleri, bir çizge yapısı olarak temsil edilir, burada her düğüm bir eklem ve kenarlar eklemler arasındaki uzaysal ilişkileri temsil eder.

Uzaysal çizge evrişimi her düğümün özelliklerini, komşu düğümleriyle birleştirerek hesaplar. Zamansal evrişimler ise iskelet verisinin zaman boyutundaki özelliklerini yakalamak için kullanılır. Şekil 17’de gösterildiği gibi $GFE_{1,2,3}$ modüllerinde bulunan ST-GC blokları üç sayı ile temsil edilmektedir. Bu sayılar sırasıyla giriş kanal sayısını, çıkış kanal sayısını ve adım boyutu değerini ifade eder. Bu bloklar 9 zamansal çekirdek boyutu kullanmaktadır. Ayrıca, her ST-GC bloğundan sonra özelliklerin aşırı uyuma yol açmaması için %50 olasılıkla rasgele elemeye tabi tutulmaktadır. B5 ve B8 bloklarının adım boyutu değerleri, havuzlama işlemi amacıyla 2 olarak ayarlanmıştır. Bu yapı sayesinde, özellik haritaları düşük seviyeli özellikler içeren haritalardan, daha yüksek seviyeli özellikler içeren haritalara kademeli olarak gelişmektedir. $GFE_{1,2}$ modüllerinden sonra, özellik haritaları $AAMG_{1,2}$ modüllerine iletilerek her eylem için en uygun eklem ilişkilerini temsil eden bitişiklik matrisleri elde edilir. GFE_3 modülü ile elde edilen özellik haritası eylemleri sınıflandırmak için sınıflandırma modülüne iletilmektedir.

Dikkat bitişiklik matrisi üretici modülü

AGMS-GCN mimarisi, iki dikkat bitişiklik matrisi üretici modülü $AAMG_{1,2}$ bulunmaktadır. $AAMG_{1,2}$ modülleri ile $GFE_{1,2}$ özellik çıkarıcı modüllerinden gelen özellik haritalarına dikkat mekanizması uygulanarak her eylem için en uygun eklem ilişkilerini temsil eden dikkat bitişiklik matrisleri elde edilmektedir. $AAMG_{1,2}$ modüllerinin genel mimarisi Şekil 18’de gösterilmiştir.



Şekil 18. AAMG Modülünün genel mimarisi

AAMG_{1,2} modüllerinde, öncelikle geçici ST-GC blokları uygulanarak özellik haritalarının daha üst düzey temsillere dönüştürülmesi sağlanır. Daha sonra, 1×1 evrişim katmanı (Conv 1×1) uygulanarak özellik haritasının boyutları, sınıf sayısına uyacak şekilde düzenlenir ve ardından toplu normalleştirme işlem uygulanır. Normalleştirme işlemi ile her özelliğin dağılımını ortalaması 0 ve standart sapmasını 1 olacak şekilde ayarlar. Bu işlem, özelliklerin benzer bir ölçekte olmasını sağlar ve bu da modelin daha hızlı ve daha tutarlı bir şekilde öğrenmesine yardımcı olur. Bu normalleştirilmiş özellik haritası daha sonra, uzaysal bilgileri korurken boyutunu küçültmek için zamansal boyut boyunca GAP'a tabi tutulur ve bu da yoğunlaştırılmış bir özellik gösterimiyle sonuçlanır. Bu işlem sonucunda her özellik haritası tek bir vektöre indirgenir. Daha sonra, özellikleri daha da iyileştirmek ve boyutlarını istenen dikkat tabanlı bitişiklik matrisiyle hizalamak için başka bir evrişimsel katmandan ve toplu normalleştirme katmanından geçirilir. Özellik haritasında, özelliklerin doğrusal olmamasını ve pozitifliğini sağlamak için bir ReLU aktivasyon fonksiyonu kullanılarak yeniden düzenlenir. Yöntemimizdeki kritik bir adım, özellik haritasının ortalama değerine dayalı olarak bitişiklik matrisi içindeki dikkat ağırlıklarının dinamik olarak ayarlanmasını içermektedir. Özellik haritasının ortalama değerine göre bir eşik değeri hesaplayarak, bu eşik değerinin altına düşen tensördeki değerler negatif olarak ayarlanır. Bu işlem bir tür adaptif eşikleme olarak düşünülebilir. Böylece, ortalama eşik altına düşen özellikler baskılanırken, daha önemli özellikler vurgulanır. Bu karşılaştırmalı etki sayesinde, önemli bağlantıları içeren bir dikkat avantajı oluşturulur. Bu süreç, özellik haritasındaki farklı bölgelerin uzaysal ilişkilerini ve önem derecelerini içeren dikkat tabanlı bitişiklik matrislerini sonuçlandırır. Elde edilen bitişiklik matrisleri, sonraki ST-GC blokları için kritik bir bileşen olarak görev yapar. Bu matrisler, evrişim işlemlerini verilerdeki daha bilgilendirici bölgelere ve ilişkilere odaklanmasına rehberlik eder. Bu dikkat tabanlı bitişiklik matrislerinden yararlanarak, yöntemimiz özellik çıkarım sürecini geliştirir ve daha ince, ayırt edici uzaysal-zamansal örüntüleri yakalar. Bu sayede, eylem tanıma performansının artırılmasına önemli bir katkı sağlanır.

Sınıflandırma modülü

AAMG₁ modülünde üretilen dikkat bitişiklik matrisleri GFE₂ modülüne aktarılır. İkinci özellik çıkarıcı modülü birinci özellik çıkarıcı modülünden gelen özellik haritasını hem sabit bitişiklik matrislerini hem de dikkat bitişiklik matrisleri ile evrişim işlemlerine tabi tutarak özellikleri toplamaktadır. GFE₂ modülünde elde edilen özellik haritası AAMG₂ modülüne aktarılır ve dikkat bitişiklik matrisleri güncellenir. Bu güncellemeden sonra, özellik haritası tekrar GFE₃ modülünde özellik haritası evrişim işlemlerine tabi tutulur ve son özellik haritası elde edilir. Son özellik haritası sınıflandırma modülüne iletilir. Sınıflandırma modülünde, özellik haritası GAP, tam bağlı doğrusal katman ve Softmax fonksiyonundan geçirilerek sınıflandırma olasılıkları ve kayıp oranı elde edilir. Son olarak, elde edilen kayıp değeri, geri yayılım sürecinde kullanılarak ağırlıklar güncellenmektedir. Geri yayılım, çıkış katmanında hesaplanan kayıp ile başlayarak bu bilginin zincir kuralı aracılığıyla ağırlık önceki katmanlarına doğru iletilmesini sağlar. Bu süreç, ağırlık daha anlamlı özellikler öğrenmesine ve sınıflandırma performansının artmasına katkı sağlamaktadır.

Deneyisel ayarlar

Deneyisel çalışmalar RTX 3090 GPU üzerinde PyTorch çerçevesi kullanılarak gerçekleştirilmiştir. Bu çalışmada kullanılan omurga Yan *et al.* (2018), Shi *et al.* (2019b) ve Shiraki *et al.* (2020) model yapılarından esinlenerek geliştirilmiştir. Modellerin eğitiminde, 0,9'lük bir Nesterov momentumuna ve 0,0004'lük bir ağırlık azalmasına (weight decay) sahip stokastik gradyan iniş (Stochastic Gradient Descent (SGD) algoritması kullanılmıştır. SGD, modelin ağırlıklarını her eğitim örneği için gradyan hesaplamalarına dayalı olarak günceller. Nesterov momentumu, gradyanların geçmiş değerlerine dayalı olarak güncellemeleri hızlandırarak, daha hızlı ve daha kararlı bir öğrenme süreci sağlar. Ağırlık azalması ise modelin aşırı uyum sağlamasını önlemeye yardımcı olur, böylece genel performansı artırır ve modelin daha iyi genelleme yapmasını sağlar. Bu kombinasyon, modelin öğrenme sürecini dengeli ve etkili bir şekilde iyileştirilmesini hedefler. Gradyanları geri yaymak için kayıp fonksiyonu olarak çapraz entropi (cross-entropy) kullanılmaktadır. Çapraz entropi kayıp fonksiyonu, sınıflandırma problemlerinde, özellikle çoklu sınıf problemlerinde, modelin tahmin ettiği sınıf olasılıkları ile gerçek sınıflar arasındaki farkı ölçer. Bu fonksiyon, modelin tahminlerinin doğruluğunu artırmaya yönelik gradyan hesaplamalarını yönlendirir ve genellikle modelin öğrenme sürecini daha etkili ve hızlı hale getirir. Eğitim prosedürünü daha kararlı hale getirmek amacıyla, ilk beş eğitim döngüsünde ısınma stratejisi (He *et al.* 2016) uygulanmıştır. Bu strateji, öğrenme oranını eğitim sürecinin başlangıcında düşük bir seviyeden başlayıp, belirli bir süre zarfında kademeli olarak artırarak modelin daha hızlı ve stabil bir şekilde öğrenmesini sağlar.

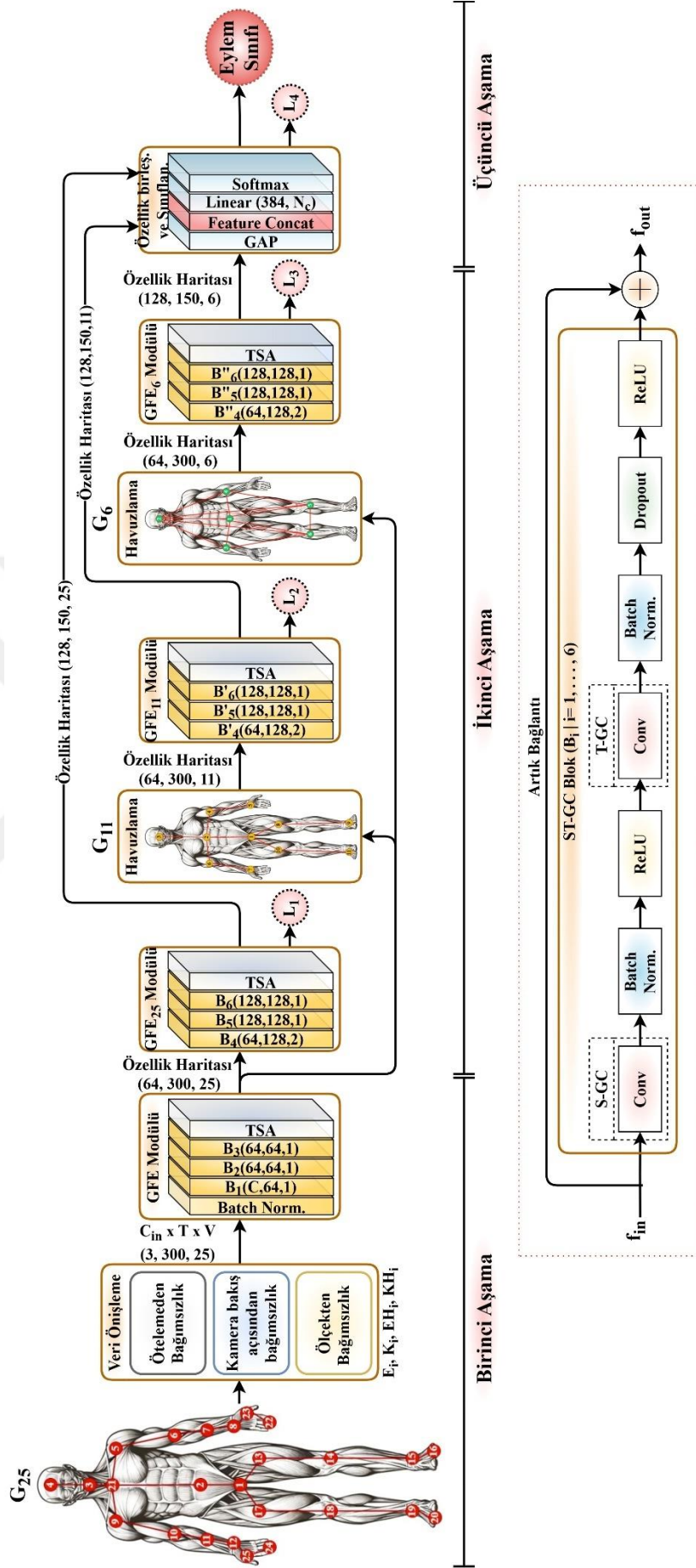
Isınma stratejisi, özellikle büyük ve karmaşık ağlarda, eğitim sürecinin daha dengeli ve etkili olmasına katkıda bulunur. Öğrenme oranını, maksimum öğrenme oranı 0.1 ve minimum öğrenme oranı 0.0001 olacak şekilde, kosinüs tavlama (Loshchilov and Hutter 2017) yöntemi ile azalacak şekilde ayarlanmıştır. Kosinüs tavlama, öğrenme oranını eğitim süreci boyunca, bir kosinüs dalgası şeklinde kademeli olarak azaltarak modelin daha iyi bir genel performans sergilemesini sağlar. Bu yöntem, öğrenme oranının başlangıçta yüksek olmasını ve eğitim sürecinin ilerleyen aşamalarında yavaşça azalmasını sağlayarak, modelin lokal minimumlardan kaçınmasını ve daha iyi bir genel sonuç elde etmesini destekler. NTU-60 ve NTU-120 veri kümeleri için toplu işlem boyutu 64, N-UCLA veri kümesi için ise 16 olarak ayarlanmıştır. Adil karşılaştırma için hiçbir veri artırma tekniğinin kullanılmamıştır. AGMS-GCN mimarisine ait tüm deneysel çalışmalarda yukarıda anlatılan yapılandırmalar kullanılmıştır.

FCSA-GCN Mimarisi

Mimari yapıya genel bakış

AGSM-GCN mimarisinde, farklı ölçekteki çizge yapıları olan G_6 , G_{11} ve G_{25} ayrı ayrı kullanmak yerine, bu yapıları tek bir akışta inceden-kabaya yaklaşımla birleştirmenin hem uzaysal-zamansal bağımlılıkları etkili bir şekilde yakalamayı sağlayacağı, hem de hesaplama maliyetlerini düşüreceği düşüncesi ile inceden-kabaya öz-dikkat çizge evrişimli ağ yapısı olan FCSA-GCN mimarisi geliştirilmiştir. FCSA-GCN genel mimarisi Şekil 19'da sunulmuştur. Özellik çıkarma modüllerindeki her ST-GC bloğu (B_i) giriş özellikleri, çıkış kanalları ve adım değeri olmak üzere üç parametre ile tanımlanmaktadır. Zamansal özellikler, zamansal öz-dikkat kullanılarak iyileştirilmektedir. Özellik haritaları giriş özellikleri, çerçeveler ve eklem sayılarına karşılık gelen üç değerle temsil edilmektedir. Kayıp değerleri L_1 , L_2 , L_3 ve L_4 olarak gösterilmektedir.

FCSA-GCN mimarisinde, insan iskeletinin uzaysal ve zamansal sınırlamalarını ele almak için çok ölçekli ve zamansal öz-dikkat mekanizmasına dayalı inceden-kabaya özellik çıkarma stratejisi benimsenmiştir. FCSA-GCN model yapısı üç ana aşamadan oluşmaktadır.



Şekil 19. FCSA-GCN'nin genel mimarisi B_i ile temsil edilen her ST-GC bloğunun iç yapısı, şeklin alt kısmında gösterilmiştir

İlk aşamada, ham iskelet verileri etkili bir ön işleme adımından geçer. Bu önemlidir çünkü aynı eylemi gerçekleştiren iki bireyin, konumlarındaki veya açılarındaki değişiklikler nedeniyle farklı eklem koordinatları olabilir. Bu tür tutarsızlıklar, modelin uzaysal yapıları doğru bir şekilde temsil etme ve etkili bir şekilde genelleme yapma yeteneğini sınırlamaktadır. Ön işleme adımı, iskelet verilerini ötelemeye, kamera bakış açısına ve ölçeğe göre değişmez hale getirerek farklı bireylerden veya senaryolardan kaynaklanan farklılıkları ortadan kaldırmayı amaçlamaktadır. Ayrıca, iskelet verilerini normalleştirmek ve tutarlı bir yapıya dönüştürmek, modelin uzaysal ve zamansal ilişkileri daha doğru bir şekilde öğrenmesini sağlar. Bu işlem, insan vücudunun doğal uzaysal yapısını etkili bir şekilde temsil eder ve özellik çıkarma sırasında kritik ayrıntıların kaybolmasını önler. Her eylem örneği için, ön işleme adımının çıktısı (3, 300, 25) şeklindeki bir tensördür. Burada, 3 giriş özelliklerini yani eklemlerin (x, y, z) koordinatlarını, 300 çerçeve sayısını ve 25 eklem sayısını temsil etmektedir. Ön işleme adımından sonra uzaysal ve zamansal ilişkiler çizge özellik çıkarıcı GFE modülü kullanılarak iskelet verilerinden elde edilmektedir. GFE modülü, ardışık ST-GC bloklarından oluşur ve 25 eklemlili bir çizge yapısı (G_{25}) kullanarak özellik çıkarma işlemini gerçekleştirir. Ancak, ST-GC blokları zamansal ilişkileri modellerken çekirdek boyutu tarafından tanımlanan sınırlı bir zamansal komşuluk içinde çalışır. Bu, onları uzun vadeli zamansal bağımlılıkları yakalamak için yetersiz hale getirir. Bu sınırlamanın üstesinden gelmek için, özellik haritasına bir zamansal öz-dikkat mekanizması uygulanır. Bu mekanizma, zamansal boyutta aynı eklem noktalarının dinamik değişimlerini yakalar ve zaman içindeki kritik eylem kalıplarını vurgular. Bu, modelin uzun vadeli zamansal hareket bağımlılıklarını daha etkili bir şekilde öğrenmesini sağlar. GFE modülüyle elde edilen özellik haritası (64, 300, 25) boyutlarındadır.

İkinci aşamada, önerilen mimari inceden-kabaya yaklaşımına dayalı olarak iskelet yapısının uzaysal karmaşıklığını azaltarak üç farklı dal boyunca özellik çıkarma sürecini sürdürür. İlk dalda, özellik haritası GFE_{25} modülü ve 25 eklemlili bir çizge yapısı kullanılarak yüksek seviyeli özellikler içeren bir forma dönüştürülür. Bu modülün çıktısı (128, 150, 25) boyutlarına sahip ince ayrıntıları içeren bir özellik haritasıdır. İkinci ve üçüncü dallarda, GFE modülünden elde edilen özellik haritasına, sırasıyla 11 (G_{11}) ve 6 (G_6) eklemlili çizge yapılarının oluşma şekillerine dayalı olarak havuzlanır ve yeni özellik haritaları oluşturulur. Bu havuzlama işlemleri, insan vücudunun koordineli hareketlerini daha etkili bir şekilde modellerken uzaysal sınırlamaları aşmayı amaçlar. Daha sonra bu özellik haritaları, yüksek seviyeli özellikleri çıkarmak için G_{11} ve G_6 çizge yapıları kullanılarak GFE_{11} ve GFE_6 modüllerinde işlenir. Bu modüllerin çıktıları, sırasıyla (128, 150, 11) ve (128, 150, 6) boyutlarında kaba ayrıntılar içeren özellik haritalarıdır. Her üç modülde de özellik kanallarının sayısı 64'ten 128'e çıkarılır. Ayrıca, bu aşamada çerçeve sayısı 300'den 150'ye düşürülür ve böylece modelin zamansal

boyutta daha öz ve anlamlı özellikler öğrenmesi ve hesaplama karmaşıklığının optimize edilmesi sağlanır. Kritik zamansal ayrıntıları daha da vurgulamak için her ölçekten çıkarılan özelliklere zamansal öz-dikkat mekanizması uygulanır. Bu inceden-kabaya yaklaşım, farklı ölçeklerde eklem ilişkilerini modelleyerek iskelet yapısının dinamik doğasını anlamayı amaçlamaktadır. Ayrıca, öğrenme sürecini desteklemek için her ölçek için ayrı kayıp değerleri elde edilir.

Üçüncü aşamada, üç çizge ölçeği G_{25} , G_{11} ve G_6 'dan çıkarılan özellik haritaları kapsamlı bir gösterim oluşturmak üzere birleştirilir. Bu birleşik özellik haritası, birden fazla ölçek boyunca uzaysal-zamansal bilgileri kapsülleyerek modelin karmaşık eylemleri ayırt etme yeteneğini artırır. Birleştirilmiş özellikler, küresel ortalama havuzlama, doğrusal bir katman ve eylem sınıflandırması için bir softmax katmanı dahil olmak üzere bir dizi tam bağlı katmandan geçirilir. Bu sınıflandırma süreci, öğrenilen özellikleri ayrı eylem sınıflarına eşleyerek nihai çıktıyı üretir. Tüm FCSA-GCN mimarisi uçtan-uca eğitilebilir bir çerçeve olarak PyTorch kütüphanesi ile inşa edilmiştir.

İnceden-kabaya özellik çıkarma stratejisi

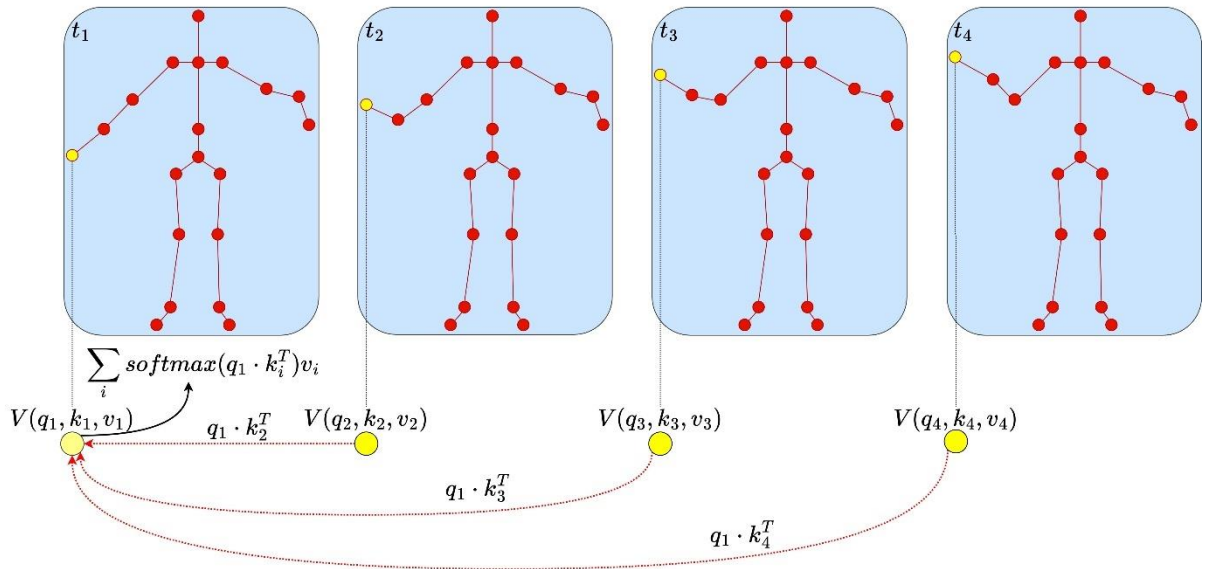
Şekil 14'de gösterildiği gibi, özellik çıkarma süreci inceden-kabaya doğru ilerleyen şekilde üç farklı çizge yapısıyla yapılmaktadır. Bu çizge yapıları, insan vücudunun farklı ölçeklerdeki temsillerini sağlar. G_{25} çizgesi 25 eklem noktasından oluşan temel iskelet yapısını ifade ederken, G_{11} çizgesi insan vücudunu 11 ana bölgeye ayırarak eklemler arasında daha üst düzey bir temsil sunar. G_6 çizgesi ise vücudu 6 daha genel bölgeye ayırarak iskelet yapısını en soyut şekilde temsil etmektedir. Bu yapı, eklemler arası bağıntıları daha geniş bir perspektiften değerlendirmeye olanak tanımaktadır.

FCSA-GCN model yapısında, ilk olarak G_{25} çizgesi kullanılarak GFE modülünde bulunan ST-GC blokları ile başlangıç seviyesinde özellik çıkarımı yapılmaktadır. Bu aşamada, iskelet verisi doğrudan işlenerek her eklem zaman içindeki değişimi ve mekânsal ilişkileri öğrenilir. Daha sonra, elde edilen özellik haritası G_{11} ve G_6 çizgelerinin oluşturulma mantığına göre havuzlanarak yeni ölçeklerde özellik haritaları oluşturulur. G_{11} ölçeğinde, 25 eklemde gelen bilgiler 11 bölgeye indirgenerek vücudun daha bütünsel hareket ilişkileri yakalanırken, G_6 ölçeğinde ise bu süreç daha da genelleştirilerek geniş kapsamlı hareket analizleri yapılmaktadır. Havuzlama işlemi sonrasında her bir ölçeğe karşılık gelen özellik haritası, ilgili çizge yapısına uygun şekilde ST-GC bloklarıyla tekrar işlenir. Bu adım, her ölçeğin kendi uzaysal ve zamansal bağıntılarını daha iyi temsil etmesine olanak tanır. Son aşamada ise, farklı ölçeklerden elde edilen özellik haritaları birleştirilerek nihai temsil vektörü oluşturulur. Bu yapı sayesinde hem eklem seviyesindeki hem de bölgesel hareket ilişkileri aynı anda göz önünde

bulundurularak kapsamlı bir analiz yapılır. Böylece model, farklı ölçeklerdeki bağıntıları etkin bir şekilde değerlendirerek daha güvenilir ve sağlam sonuçlar üretmektedir.

Zamansal öz-dikkat

ST-GCN modellerinde zamansa evrişim $1 \times K_t$ boyutlu bir çekirdek ile çalışır ve her işlemde K_t çerçeve boyunca özellik çıkarımı yapılır. Bu zamansal çekirdek boyutunun sınırlılığı, özellikle uzun menzilli bağımlılıkları yakalamakta yetersiz kalmaktadır. Geleneksel K_t -boyutlu evrişimler, sadece sabit bir çerçeve aralığındaki (örneğin $K_t = 5$ veya $K_t = 9$) ilişkileri öğrenebilir. Ancak insan hareketleri, farklı zaman ölçeklerinde değişen karmaşık bağımlılıklar içerebilir. Bu nedenle, FC-SA-GCN mimarisinde eklemlerin zamansal bağlamını daha etkili bir şekilde modellemek için zamansal öz-dikkat (Temporal Self-Attention (TSA)) mekanizması kullanılmaktadır. TSA ile zenginleştirilmiş zamansal evrişimler, zaman boyutundaki hem kısa hem de uzun vadeli bağımlılıkları etkili bir şekilde modelleyerek hareketlerin statik ve dinamik bağlamını daha iyi öğrenilmesini sağlamaktadır. TSA, Transformer modelinde kullanılan öz-dikkat mekanizmasının zaman temelli verilere uyarlanmış bir versiyonudur. Transformer modellerinde öz-dikkat, giriş dizisindeki öğelerin birbiriyle olan ilişkilerini öğrenerek bağlam bilgisi oluşturmayı hedeflerken, TSA bu mekanizmayı zaman eksenine taşır. Şekil 20’de bir el eklemine (t_1 ’de sarı ile vurgulanan nokta) diğer zaman dilimindeki (t_2, t_3, t_4) durumlarıyla olan dikkate dayalı ilişkisi gösterilmektedir.

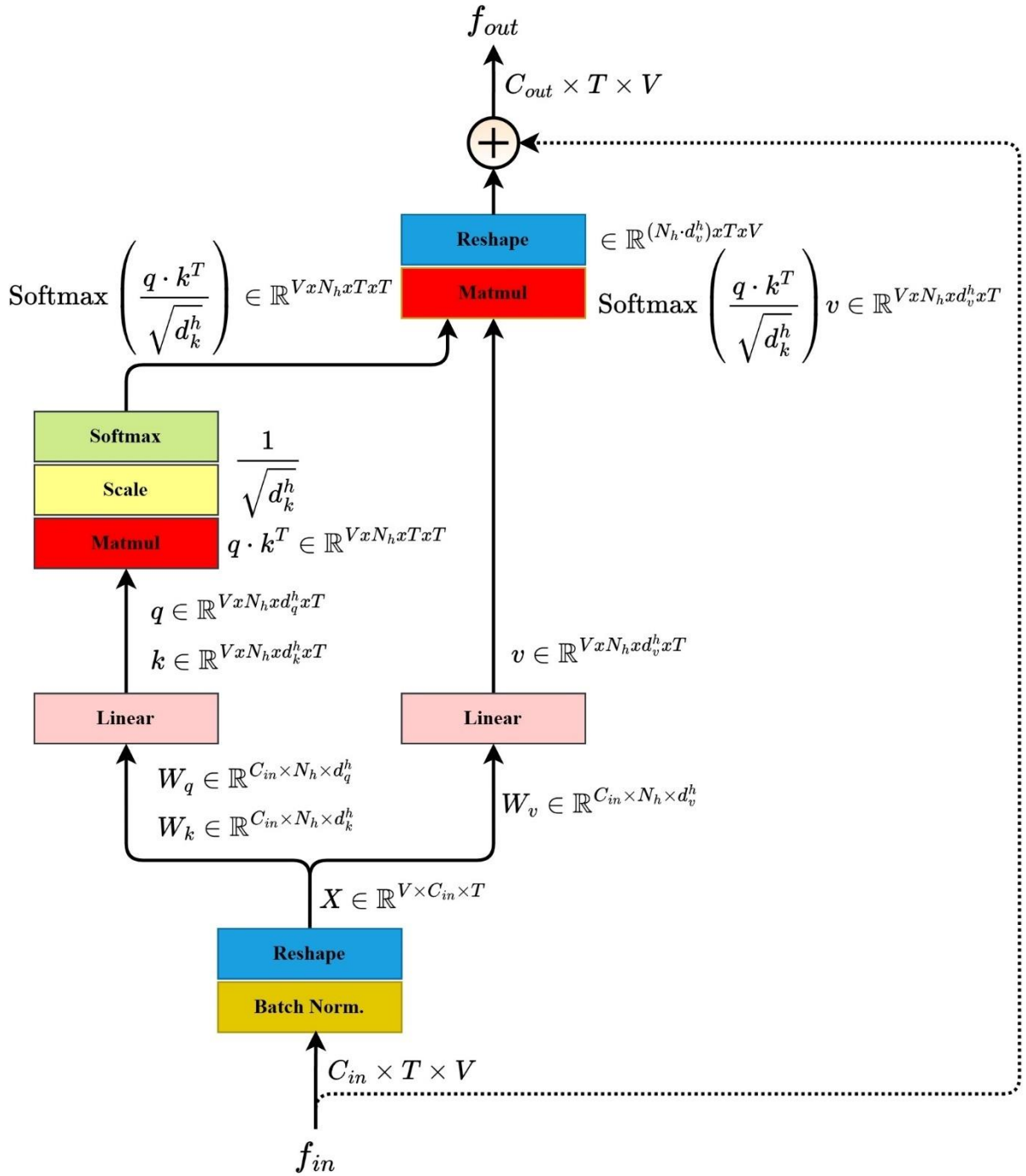


Şekil 20. Zamansal öz-dikkat mekanizmasının iskelet çerçeveler boyunca çalışması

Şekil 20’deki kırmızı kesikli çizgiler, t_1 ’deki eklem ile diğer zaman dilimindeki aynı eklem arasında kurulan bağları göstermektedir. TSA mekanizması, bu ilişkileri öğrenmek için her bir eklem noktasını sorgu (q), anahtar (k) ve değer (v) matrisleri ile temsil eder. t_1 ’deki el

ekleminin diğer zaman dilimlerindeki durumlarıyla olan bağımlılığı, $q_1 \cdot k_i^T$ formülüyle dikkat skorları hesaplanarak belirlenir. Bu skorlar, softmax fonksiyonu ile normalleştirip v_i ile ağırlıklandırılarak zaman eksenindeki her bir çerçevede el ekleminin kendi hareketine dair bir bağlam oluşturur. Şekil 20'de örnek olarak yalnızca dört çerçeve verilmiştir; ancak TSA mekanizması, eylemin tüm çerçevelerini kullanarak eklem noktalarının zaman eksenindeki dinamik değişimlerini yakalar. Bu, standart ST-GCN yöntemleri ile çıkarılamayan uzak menzilli ayırt edici özelliklerin öğrenilmesine olanak tanır ve eylem tanıma görevlerinde performansı artırmaktadır.

Şekil 21'de Bello *et al.* (2019) ve Plizzari *et al.* (2021) tarafından önerilen yöntemlere dayalı olarak iskelet dizilerinde TSA mekanizmasının uygulanmasını göstermektedir. Giriş tensörü f_{in} , $(C_{in} \times T \times V)$ olarak yapılandırılmıştır; burada C_{in} giriş özelliklerini, T çerçeveleri ve V eklemleri ifade eder. Giriş tensörü önce bir toplu normalleştirme katmanı ile işlenir. Daha sonra, tensör $X \in \mathbb{R}^{V \times C_{in} \times T}$ matrisine yeniden düzenlenir ve V boyutu öne alınır. Bu yeniden düzenleme, her eklem için zamansal boyut boyunca ayrı işlem yapmayı mümkün kılar.



Şekil 21. Zamansal öz-dikkat mekanizmasının iskelet dizilerine uygulanması

X matrisine çok başlı dikkat (Multi-Head Attention (MHA)) uygulanır. Böylece dikkat hesaplamalarının birden fazla başlık (h) kullanılarak paralel olarak gerçekleştirilmesine olanak tanır. Her kafa, modelin giriş verilerinin farklı yönlerini aynı anda yakalamasını sağlayan ayrı bir dikkatli öğrenme süreci gerçekleştirir. Her kafa için öz-dikkat mekanizmasının çıktısı, Denklem (13) kullanılarak hesaplanır.

$$Out_h(X) = Softmax \left(\frac{(XW_q)(XW_k)^T}{\sqrt{d_k^h}} \right) (XW_v) \quad (13)$$

Denklem (13)'de $W_q \in \mathbb{R}^{C_{in} \times N_h \times d_q^h}$, $W_k \in \mathbb{R}^{C_{in} \times N_h \times d_k^h}$ ve $W_v \in \mathbb{R}^{C_{in} \times N_h \times d_v^h}$ sırasıyla X 'i sorgulara ($q = XW_q$), anahtarlara ($k = XW_k$) ve değerlere ($v = XW_v$) dönüştüren öğrenilebilir doğrusal dönüşüm ağırlık matrisleridir. Ağırlık matrislerindeki N_h dikkat başlıklarının sayısını, d_q^h , d_k^h ve d_v^h ise her başlık için sorgu, anahtar ve değer matrislerinin derinliğini ifade etmektedir. Sorgu ve anahtar matrisleri çarpılarak dikkat skorları ($q \cdot k^T$) elde edilir, ardından bu skorlar ölçeklendirilir ve softmax işlemi ile normalleştirilir. Sonuç olarak, bu işlemlerden elde edilen çıktı, $(V \times N_h \times T \times T)$ boyutlarına sahip bir dikkat matrisidir. Bu matris, aynı eklem noktalarının farklı zaman adımlarında birbirleriyle ne ölçüde ilişkili olduğunu gösteren skaler değerler içermektedir. Dikkat matrisi, zamansal boyut boyunca bağlamsal bilgileri zenginleştiren bir tensör elde etmek amacıyla değer matrisi ile çarpılır. Farklı kafaların çıktıları Denklem (14) kullanılarak birleştirilir.

$$MHA(X) = Concat(Out_1, \dots, Out_{N_h})W^o \quad (14)$$

Denklem (12)'de, $Out_1, \dots, Out_h$ her dikkat başlığından elde edilen çıktılarıdır ve W^o başlıkların çıktılarını birleştiren öğrenilebilir bir doğrusal dönüşüm ağırlık matrisidir. Son adımda, oluşturulan tensör yeniden şekillendirilir ve girişten gelen f_{in} ile toplanarak f_{out} üretilir. Bu işlem hem giriş bilgisinin korunmasını hem de zaman boyunca öğrenilen bağlamsal bilgilerin eklenmesini sağlar. TSA mekanizması, zaman çizelgesi boyunca her eklem için zamansal varyasyonların etkili bir şekilde öğrenilmesini mümkün kılmaktadır.

Toplam kayıp değerlerinin DWA ile optimizasyonu

Birçok çoklu görev öğrenme ağında, birden fazla görevi aynı anda eğitmek zordur çünkü görevlerin kayıp fonksiyonları arasındaki dengeyi bulmak gereklidir. Bu sorunun üstesinden gelmek için literatürde bazı çalışmalar yapılmıştır (Chen *et al.* 2018; Kendall *et al.* 2018; Liu *et al.* 2019). Özellikle Liu *et al.* (2019) tarafından önerilen dinamik ağırlık ortalaması (Dynamic Weight Averaging (DWA)) yöntemi, derin öğrenme modellerinde çoklu görev öğrenimi bağlamında kullanılır ve her bir görevin kaybının göreceli zorluğuna dayalı olarak ağırlıkları dinamik olarak ayarlar. Böylece, farklı zorluk seviyelerine sahip görevler optimize edilirken daha zor görevlere daha fazla odaklanılır. DWA'nın temel mantığı, her görevin kayıplarının zaman içindeki değişim oranlarına göre ağırlıkları belirlemektir. Başlangıçta kayıplar yakın olsa da bazı görevler daha hızlı optimize edilirken diğerleri daha uzun sürebilir. DWA bu

farkları tespit ederek ağırlıkları sürekli günceller ve tüm görevlerin dengeli bir şekilde optimize edilmesini sağlar. Matematiksel olarak, kayıp fonksiyonlarının önceki eğitim döngülerindeki değişim oranlarına dayalı ağırlık güncellemeleriyle bu işlem gerçekleştirilir.

Önerilen FCSEA-GCN mimarisinde, eğitim sürecinde farklı ölçeklerdeki G_{25} , G_{11} ve G_6 çizgelerinden ayrı ayrı özellik haritaları üretilmektedir. Bu haritaların birleştirilmesiyle ortak bir özellik haritası oluşturulmaktadır. Her bir özellik haritasından bağımsız olarak eylem sınıfları tahmin edilmekte ve kayıp değerleri (Şekil 19'daki L_1 , L_2 , L_3 ve L_4) hesaplanmaktadır. DWA tipik olarak birden fazla görevi içeren senaryolar için tasarlanmış olsa da tek bir görev içindeki birden fazla kayıp fonksiyonunun katkılarını dengelemek için de etkili bir şekilde uyarlanabilir. FCSEA-GCN modelinde, DWA bağımsız görevleri optimize etmek için değil, farklı ölçeklerdeki özellik haritalarından türetilen kayıpların ağırlık katkılarını dinamik olarak ayarlamak için kullanılmaktadır. Bu optimizasyon süreci, farklı kayıpların etkilerini dengeli bir şekilde hesaba katarak daha iyi bir yakınsama ve model performansı sağlar. Birleştirilmiş özellik haritasının kaybı olan L_4 model performansı üzerindeki kritik önemi nedeniyle DWA'dan hariç tutulur ve doğrudan toplam kayba eklenir. DWA yalnızca kayıp fonksiyonlarının ağırlıklarını dinamik olarak günceller. Bu ağırlıklar, eğitim süreci sırasında modelin öğrenmesini yönlendirmek için kullanılır, ancak modelin kendisi içinde eğitilebilir parametreler olarak tanımlanmazlar. Bu nedenle, eğitilebilir parametrelerin toplam sayısı ve modelin mimarisi değişmeden kalır. Benzer şekilde, hesaplama karmaşıklığının önemli bir ölçüsü olan FLOP değeri de DWA'dan etkilenmez. DWA eğitim süresini biraz artırır, ancak bu model yapısındaki değişikliklerden ziyade eğitim sırasında gerçekleştirilen ek işlemlerden kaynaklanır. DWA dört adımdan meydana gelmektedir.

Adım 1- Kayıp oranlarının hesaplanması: Her bir kayıp fonksiyonu için, önceki iki eğitim döngüsündeki kayıp değerleri kullanılarak kayıp oranları hesaplanır. Bu oranlar, kayıpların zaman içindeki değişimini yansıtır. Her bir kayıp fonksiyonu için kayıp oranları Denklem (15) ile hesaplanır.

$$r_i(t) = \frac{L_i(t-1)}{L_i(t-2)} \quad (15)$$

Denklem (15)'teki $r_i(t)$, i . kayıp fonksiyonunun t . eğitim döngüsündeki kayıp oranını göstermektedir. $L_i(t-1)$, i . kayıp fonksiyonunun $(t-1)$. eğitim döngüsündeki değeri, $L_i(t-2)$, i . kayıp fonksiyonunun $(t-2)$. eğitim döngüsündeki değerini göstermektedir.

Adım 2- Kayıp oranlarının normalleştirilmesi: Kayıp fonksiyonlarının ölçeklerinin benzer olması, DWA'nın daha etkili çalışmasını sağlar. Bu nedenle, farklı ölçeklerdeki kayıpların etkisini dengelemek için kayıp oranları normalleştirilmesi gerekmektedir. Bu işlem,

tüm kayıp oranlarının ortalamasının hesaplanması ve her bir oranın bu ortalama bölünmesi ile gerçekleştirilir. Kayıp oranlarının normalleştirilmesi Denklem (16)'da ifade edilmiştir.

$$\bar{r}_i(t) = \frac{r_i(t)}{\frac{1}{K} \sum_{j=1}^K r_j(t)} \quad (16)$$

Denklem (16)'da $\bar{r}_i(t)$, i . kayıp fonksiyonunun t . eğitim döngüsündeki normalleştirilmiş kayıp oranını göstermektedir. K toplam kayıp fonksiyonu sayısını temsil etmektedir.

Adım 3- Ağırlıkların güncellenmesi: DWA yönteminde, her bir kayıp fonksiyonunun ağırlığı, kaybın önceki eğitim döngülerindeki değişimine göre dinamik olarak güncellenir. Ağırlıklar, normalleştirilmiş kayıp oranlarının üstel fonksiyonu alınarak hesaplanır. Üstel fonksiyon, oranlardaki küçük değişikliklerin ağırlıklar üzerinde daha belirgin bir etkisi olmasını sağlar. Bu etkinin derecesi, sıcaklık parametresi T ile kontrol edilir ve ağırlıkların ne kadar agresif değişeceğini kontrol eder. Dinamik ağırlıklar, normalize edilmiş kayıp oranlarına dayalı olarak Denklem (17) ile hesaplanır.

$$w_i(t) = \frac{K \cdot \exp\left(\frac{\bar{r}_i(t)}{T}\right)}{\sum_{j=1}^K \exp\left(\frac{\bar{r}_j(t)}{T}\right)} \quad (17)$$

Denklem (17)'deki $w_i(t)$, i . kayıp fonksiyonunun t . eğitim döngüsündeki ağırlığını, K toplam kayıp fonksiyonu sayısını göstermektedir. İlk iki eğitim döngüsünde kayıp değerleri olan $L_i(t-1)$ ve $L_i(t-2)$ mevcut olmadığından, DWA ağırlıklarını hesaplamak mümkün değildir bu yüzden tüm kayıp fonksiyonlarının ağırlıkları eşit olarak alınmaktadır ($w_i(1) = w_i(2) = w_i(3) = 1.0$).

Adım 4- Toplam kayıp değerinin hesaplanması: Toplam kaybın hesaplanması Denklem (18)'de verilmiştir. Denklem (18)'de $w_i(t)$, t . eğitim döngüsündeki i . kayıp fonksiyonun ağırlığını $L_i(t)$, t . eğitim döngüsündeki i . kayıp fonksiyonun değerini temsil eder.

$$Total Loss = \sum_{i=1}^K w_i(t) \cdot L_i(t) \quad (18)$$

DWA yöntemi, Şekil 19'da gösterilen her ölçekte özellik çıkarma sırasında elde edilen kayıp değerlerinin (L_1 , L_2 ve L_3) katkılarını eğitim süreci boyunca dinamik olarak ayarlar. Bu yaklaşım, her kaybın model üzerindeki etkisini optimize ederek dengeli öğrenmeyi sağlar. Buna karşılık, kayıp L_4 doğrudan toplam kayıp değerine eklenir ve model üzerindeki etkisi vurgulanır. Diğer kayıpların DWA ile ağırlıklandırılması, modelin diğer çıktıları dengeli bir

şekilde öğrenmesine yardımcı olmaktadır. Bu, modelin bu kritik çıktıyı daha etkili bir şekilde öğrenmesini sağlar. Toplam kayıp değeri, ağırlık geri yayılım sürecinde kullanılır.

Deneysel ayarlar

Deneysel çalışmalar, RTX 4080 16 GB GPU kullanılarak gerçekleştirilmiştir. GPU belleği ile ilgili sınırlamaları aşmak için gradyan biriktirme (gradient accumulation) yöntemi uygulanmıştır. FCSA-GCN mimari yapısı, PyTorch çerçevesi kullanılarak Yan *et al.* (2018) ve Shiraki *et al.* (2020) model yapılarından esinlenerek geliştirilmiştir. Modellerin eğitiminde, 0,9'luk bir Nesterov momentumuna ve 0,0005'lük bir ağırlık azalmasına sahip stokastik gradyan iniş algoritması kullanılmıştır. Kayıp fonksiyonu olarak çapraz entropi seçilmiş ve kayıp fonksiyonlarının ağırlıklarını dinamik olarak ayarlamak için DWA algoritması uygulanmıştır. Eğitim prosedürünü daha kararlı hale getirmek amacıyla, ilk beş dönem boyunca ısınma stratejisi (He *et al.* 2016) benimsenmiştir. Başlangıç öğrenme oranı 0,1 olarak belirlenmiş ve bu oran, eğitim süresince kosinüs tavlama (Loshchilov and Hutter 2017) yöntemiyle kademeli olarak azaltılmıştır. Adil bir karşılaştırma yapmak amacıyla hiçbir veri artırma tekniği kullanılmamıştır. FCSA-GCN mimarisine ait tüm deneysel çalışmalarda yukarıda anlatılan yapılandırmalar kullanılmıştır.

ARAŞTIRMA BULGULARI VE TARTIŞMA

Değerlendirme Protokolleri

NTU-60 veri kümesi için Shahroudy *et al.* (2016) çalışmalarında, Çapraz Denek (Cross Subject (CS)) ve Çapraz Görünüm (Cross View (CV)) olmak üzere iki değerlendirme protokolü önermişlerdir. CS değerlendirme protokolünde veri kümesi, denek numaralarına göre eğitim ve test kümelerine ayrılırken, CV değerlendirme protokolünde veri kümesi, kamera bakış açılarına göre eğitim ve test kümelerine ayrılmaktadır. NTU-120 veri kümesi için ise Liu *et al.* (2019), çalışmalarında Çapraz Denek (Cross Subject (CS)) ve Çapraz Kurulum (Cross Setup (CE)) olmak üzere iki değerlendirme protokolü önermişlerdir. CS değerlendirme protokolünde veri kümesi, denek numaralarına göre eğitim ve test kümelerine ayrılırken, CE değerlendirme protokolünde veri kümesi, kamera kurulum kimliklerine göre eğitim ve test kümelerine ayrılmaktadır.

N-UCLA veri kümesi için Wang *et al.* (2014) çalışmalarında, ilk iki kameradan alınan verileri eğitim için diğer kameradan alınan verilerin ise test için kullanılmasını önermişlerdir. Bu doktora çalışmasında deneysel çalışmalar için yukarı anlatılan değerlendirme protokolleri takip edilmiştir. Tablo 4’de, veri kümelerinin değerlendirme protokolleri doğrultusunda eğitim ve test kümelerindeki veri sayıları, eksik/bozuk veri sayıları ve toplam veri sayıları gibi bilgiler sunulmaktadır. Toplam veri miktarına eksik verilerde dahil edilmiştir.

Tablo 4. Değerlendirme Protokollerine Dayalı Olarak Veri Kümelerinin Eğitim ve Test Bölümlerine Ayrılması

Veri Kümeleri	Protokol	Eğitim Verisi	Test Verisi	Eksik Veriler	Toplam Veri
N-UCLA (Wang <i>et al.</i> 2014)	CV	1.019	463	12	1.494
NTU-60 (Shahroudy <i>et al.</i> 2016)	CS	40.091	16.487	302	56.880
	CV	37.646	18.932		
NTU-120 (Liu <i>et al.</i> 2019)	CS	63.026	50.919	535	114.480
	CE	54.468	59.447		

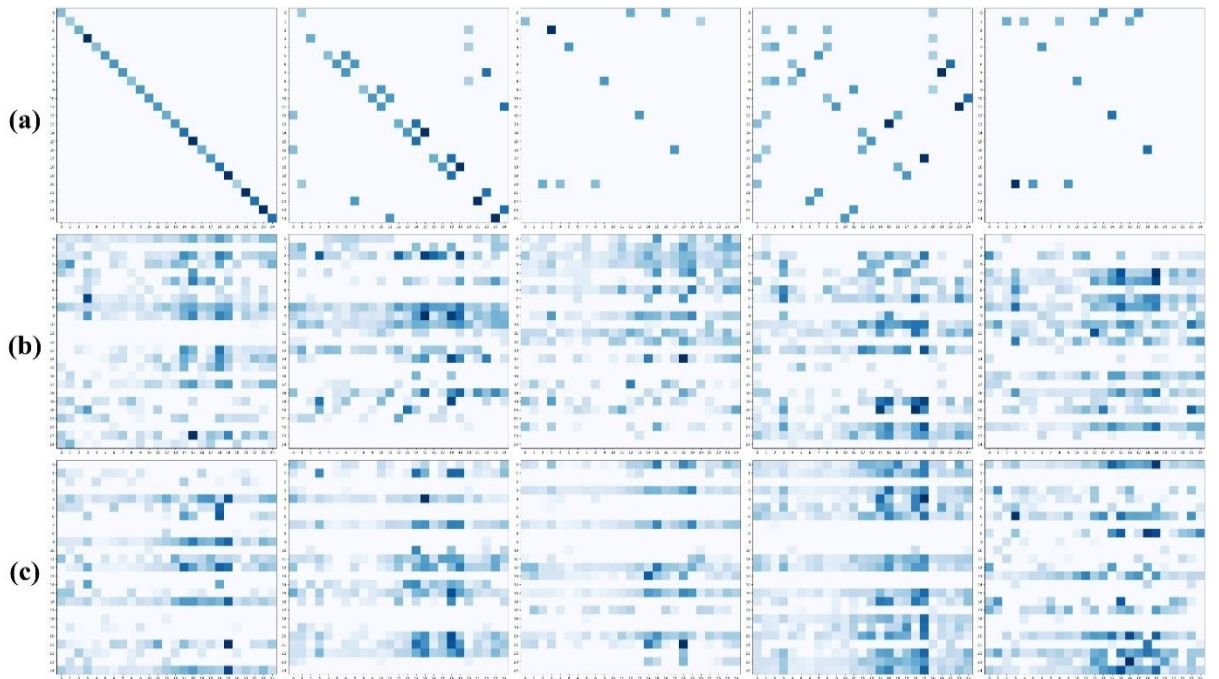
AGMS-GCN Mimarisinin Deney Sonuçları

Bu bölümde, önerilen AGMS-GCN mimarisinin farklı bileşenlerinin etkisini değerlendirmek amacıyla ayrıntılı bir ablasyon çalışması gerçekleştirilmiştir. İlk olarak yöntem tarafından eğitim aşamasında oluşturulan dikkat bitişiklik matrislerinin görselleştirilmesi ve

detaylı açıklaması gerçekleştirilmiştir. Daha sonra çoklu modalite öğrenmenin doğruluk üzerindeki etkisi analiz edilmiştir. Ardından çok ölçekli yapıların doğruluk üzerindeki etkisi araştırılmıştır. Son olarak, önerilen yöntemin performansı mevcut son teknoloji yöntemlerle karşılaştırılarak genel başarımı ortaya konmuştur.

Dikkat bitişiklik matrislerinin görselleştirilmesi

Şekil 22’de NTU-60 çapraz görünüm kıyaslaması sırasında 25 eklem noktasından oluşan G_{25} çizgesi için modelimiz tarafından kullanılan sabit bitişiklik matrislerinin bir örneğini ve $AAMG_{1,2}$ modülleri tarafından oluşturulan bitişiklik matrislerini göstermektedir. Matrislerdeki (i, j) konumundaki her bir eleman, i eklemi ile j eklemi arasındaki bağlantının ağırlığını temsil eder. Matristeki her bir elemanın mavi ölçeği, bu ağırlığın gücünü göstermektedir. Şekil 22 (a)’da normalize edilmiş sabit bitişiklik matrislerini göstermektedir. Atlama boyutu değeri 2’dir. Bitişiklik matrisleri 0, 1 ve 2 atlama boyutlarıyla fiziksel olarak bağlı olan eklemeler arasındaki bağımlılıkları temsil eden seyrek matrislerdir. Her matris, düğümler arasındaki mesafeye ve merkez düğüme olan mesafeye göre farklı kategorilere ayrılmıştır. İlk matris, 0-hop atlama mesafesinde düğümler arasındaki bağlantıları temsil eder. İkinci ve üçüncü matrisler, 1-hop atlama mesafesinde olan ve sırasıyla merkez düğüme daha yakın ve daha uzak olan düğümler arasındaki bağlantıları temsil eder. Dördüncü ve beşinci matrisler, sırasıyla merkez düğüme daha yakın ve daha uzak olan 2-hop atlama mesafesinde olan düğümler arasındaki bağlantıları temsil eder. Bu şekilde, eklemeler arasındaki ilişkiler daha ayrıntılı olarak modellenmektedir.



Şekil 22. (a) Normalize edilmiş sabit bitişiklik matrisleri, (b) $AAMG_1$ modülü tarafından üretilen bitişiklik matrisleri, (c) $AAMG_2$ modülü tarafından üretilen bitişiklik matrisleri

Şekil 22 (a) ve (b), sırasıyla AAMG₁ ve AAMG₂ modülleri tarafından üretilen bitişiklik matrislerini gösterir. Bu dikkat mekanizması tarafından öğrenilen matrisler daha yoğun ve daha az seyrek ve fiziksel olarak bağlı olmayan eklemler arasındaki bağımlılıkları da ifade eder. Bu nedenle dikkat matrisleri sayesinde, yalnızca yapısal bağımlılıklar değil aynı zamanda yapısal olmayan bağımlılıklar da modele dahil edilebilir ve iskelet dizilerini anlamada daha fazla esneklik ve bilgi sağlar. Bu öğrenilen matrislerin birbirinden farklı olması, modelin çok seviyeli anlamsal bilgi çıkarma yeteneğini artırır ve eklemler arasındaki daha karmaşık bağımlılıkların keşfetmesine olanak tanımaktadır.

Çoklu modalite öğrenmenin doğruluk üzerindeki etkisi

Model performansı üzerinde farklı girdi verisi tiplerinin etkisini incelemek için deneysel çalışmalar yürütülmüş ve sonuçlar Tablo 5’de gösterilmiştir. Tablo 5, eklem ve kemiklerin konum bilgilerinin hem CS hem de CV ayarlarında en yüksek doğruluk oranlarını sağladığını göstermektedir (%92,7 ve %96,8). Bu, konum bilgisinin eylem tanıma görevinde kritik olduğunu göstermektedir. Hareket bilgisi de yüksek doğruluk oranlarına ulaşmıştır, ancak konum bilgisinden biraz daha düşüktür. Konum ve hareket bilgilerini birleştirmenin yararlı olabileceği, ancak konum bilgisinin daha belirleyici olduğu sonucuna varılmıştır. Konum ve hareket bilgilerinin farklı ağırlarda işlenmesinin nedeni, eklem koordinatlarının uzaysal konum bilgisini, hareket bilgisinin ise zamansal hareket bilgisini içermesidir. Konum ve hareket bilgilerinin hem özellikleri hem de ölçekleri farklı olduğundan, bunları aynı ağa girdi olarak vermek öğrenme sürecini karmaşıklatacağı için makul bir yaklaşım olmayacaktır. Bu nedenle, konum ve hareket bilgilerinin ayrı ayrı ele alınması ve değerlendirilmesi öğrenme sürecini daha etkili hale getirmektedir.

Tablo 5. AGMS-GCN Mimarisinin NTU-60 Veri Kümesinde Farklı Modalitelerde Elde Ettiği Doğruluk (%) Oranlarının Karşılaştırılması

Modalite	CS	CV
Eklem (Konum, Hareket)	89,9	93,6
Kemik (Konum, Hareket)	90,3	93,3
Konum (Eklem, Kemik)	92,7	96,8
Hareket (Eklem, Kemik)	90,7	95,1

Çok ölçekli yaklaşımın doğruluk üzerindeki etkisi

Farklı ölçeklerdeki çizge yapılarının eylem tanıma performansına olan etkisinin deneysel olarak incelenmesiyle elde edilen sonuçlar Tablo 6 ve Tablo 7’de verilmiştir. Her iki tabloda da G₂₅ çizge yapısı en yüksek doğruluk oranlarına ulaşmıştır (%93,5 ve %97,5). Bu da G₂₅’in uzaysal bilgi gösterimi ve eylem tanıma için en etkili çizge yapısı olduğunu

göstermektedir. G_{11} orta düzeyde iyi performans göstererek hem konum hem de hareket verilerini temsil etmede makul bir verimlilik göstermektedir. G_6 en düşük doğruluk oranlarına sahip olup bu da uzaysal bilgiyi temsil etmede diğer çizge yapılarından daha az etkili olduğunu göstermektedir. Ancak çizge yapılarını birleştiren topluluk modelleri en yüksek doğruluk oranlarını sağlayarak bireysel çizge yapılarının performansını aşmaktadır. Bu istatistikler farklı çizge yapılarının sonuçlarını birleştirmenin eylem tanıma görevlerinde genel performansı iyileştirmek için etkili bir strateji olduğunu göstermektedir.

Tablo 6. AGMS-GCN Mimarisinin NTU-60 Veri Kümesinde Farklı Ölçekli Çizge Yapılarıyla CS Protokolünde Elde Ettiği Doğruluk Oranları (%)

Çizge Yapısı	Konum	Hareket	Konum & Hareket
G_6	88,4	86,6	89,7
G_{11}	89,9	88,4	91,0
G_{25}	92,7	90,7	93,5
$G_6 + G_{11} + G_{25}$			93,9

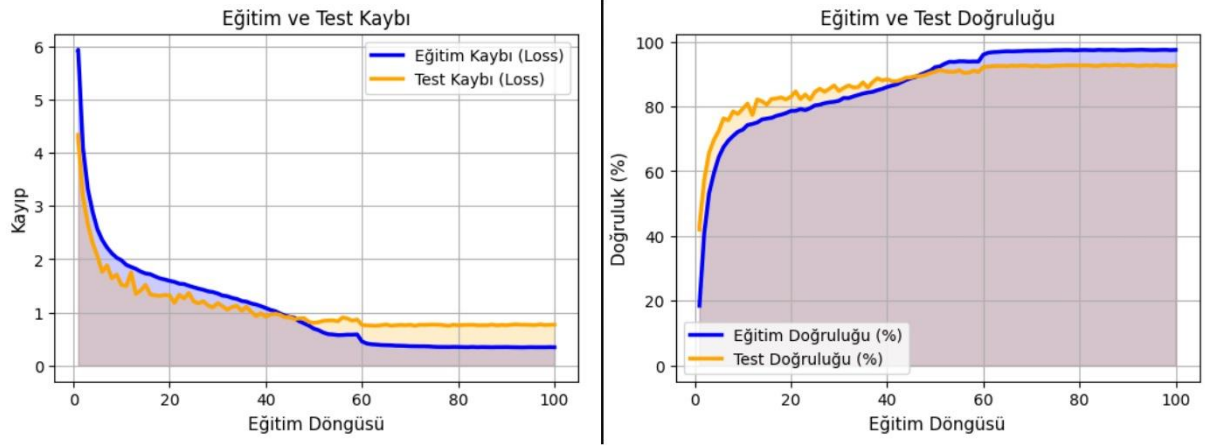
Tablo 7. AGMS-GCN Mimarisinin NTU-60 Veri Kümesinde Farklı Ölçekli Çizge Yapılarıyla CV Protokolünde Elde Ettiği Doğruluk Oranları (%)

Çizge Yapısı	Konum	Hareket	Konum & Hareket
G_6	92,1	89,7	93,0
G_{11}	93,7	92,1	94,6
G_{25}	96,8	95,1	97,5
$G_6 + G_{11} + G_{25}$			97,8

Öğrenme sürecinin görsel analizi

Derin öğrenme tabanlı derin modellerin başarısı, yalnızca son doğruluk değerleriyle değil, aynı zamanda eğitim sürecinde elde edilen performans eğilimleriyle de değerlendirilmektedir. Bu performans eğilimleri, modelin eğitim boyunca gösterdiği doğruluk ve kayıp değişimlerini, öğrenmenin sağlıklı ilerleyip ilerlemediğini, modelin genelleme yeteneğini ve aşırı öğrenme riskini ortaya koymaktadır. Bu bölümde, geliştirilen AGMS-GCN mimarisinin NTU-60 veri kümesi üzerinde CS ve CV değerlendirme protokollerine göre gerçekleştirdiği eğitim süreçleri grafiksel olarak sunulmakta ve analiz edilmektedir.

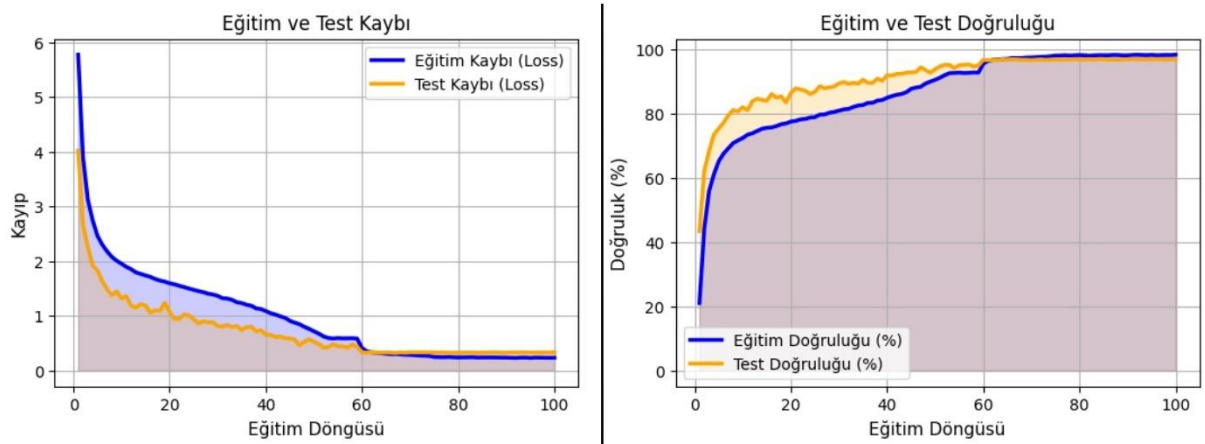
Şekil 23'te, AGMS-GCN mimarisinin nihai modelinin NTU-60 veri kümesi üzerinde CS değerlendirme protokolü kapsamında konum (eklem, kemik) modalitesinde G_{25} çizge yapıyla elde edilen eğitim sürecine ait kayıp ve doğruluk grafikleri sunulmuştur.



Şekil 23. AGMS-GCN mimarisinin NTU-60 veri kümesi üzerinde CS protokolü kapsamında konum modalitesine ait eğitim grafikleri

CS senaryosundaki eğitim süreci incelendiğinde eğitim sürecinde, eğitim kaybının düzenli olarak azaldığı ve eğitim doğruluğunun sürekli olarak arttığı gözlemlenmektedir. Eğitim sürecinde erken durdurma sabır değeri 7 olarak belirlenmiştir. Test doğruluğu ve kaybı açısından en iyi performans 78. eğitim döngüsünde elde edilmiş, bu noktadan sonra 7 ardışık eğitim döngüsü boyunca (79-85) test performansında anlamlı bir iyileşme sağlanamamıştır. Grafikte toplam 100 eğitim döngüsü yer almakta olmasına rağmen, en iyi performansın 78. eğitim döngüsünde elde edilmesi nedeniyle nihai model bu noktada kaydedilmiştir. Böylece, modelin aşırı öğrenmeye geçmeden en iyi genelleme performansına ulaştığı nokta elde edilmiştir. Eğitim sürecinin başlangıçta 100 eğitim döngüsü olarak planlanmış olması nedeniyle model eğitimin sonuna kadar takip edilmiş ve model davranışının detaylı analizi sağlanmıştır.

Şekil 24'te, AGMS-GCN mimarisinin nihai modelinin NTU-60 veri kümesi üzerinde CV değerlendirme protokolü kapsamında konum (eklem, kemik) modalitesinde G_{25} çizge yapısıyla elde edilen eğitim sürecine ait kayıp ve doğruluk grafikleri sunulmuştur.

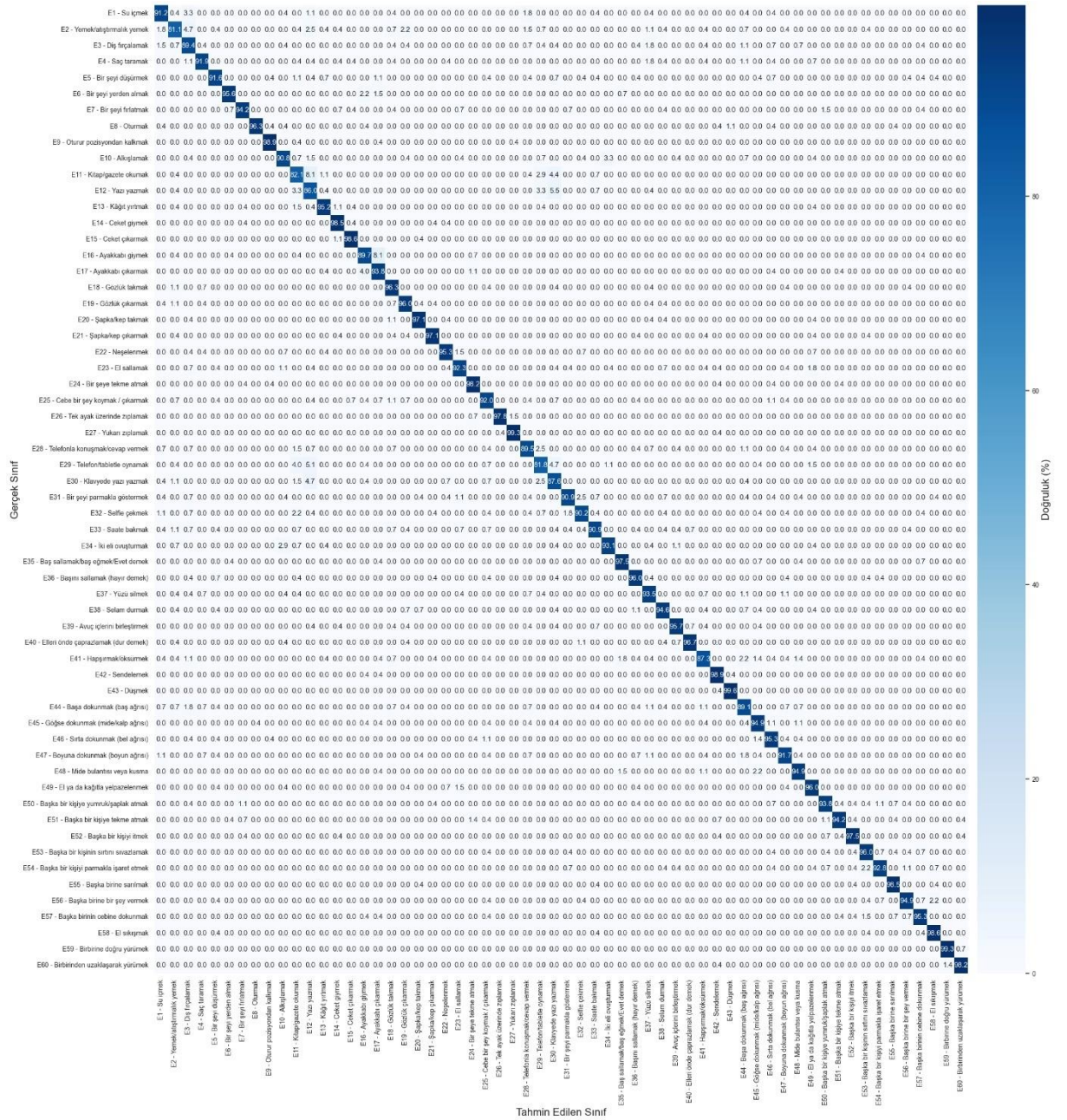


Şekil 24. AGMS-GCN mimarisinin NTU-60 veri kümesi üzerinde CV protokolü kapsamında konum modalitesine ait eğitim grafikleri

CV senaryosundaki eğitim süreci incelendiğinde eğitim kaybının sürekli azaldığı ve eğitim doğruluğunun artış gösterdiği gözlemlenmiştir. Test kaybı açısından en iyi performans 67. eğitim döngüsünde elde edilmiş, ancak bu noktadan sonra 7 ardışık eğitim döngüsü boyunca (68-74) test performansında anlamlı bir iyileşme sağlanamamıştır. Grafikte toplam 100 eğitim döngüsü yer almakta olsa da en iyi performansın 67. eğitim döngüsünde elde edilmesi nedeniyle nihai model bu noktada kaydedilmiştir. Bu durum, modelin aşırı öğrenmeye geçmeden en iyi genelleme performansına ulaştığı noktada başarıyla durdurulduğunu göstermektedir.

Mimarinin sınıf bazlı değerlendirilmesi

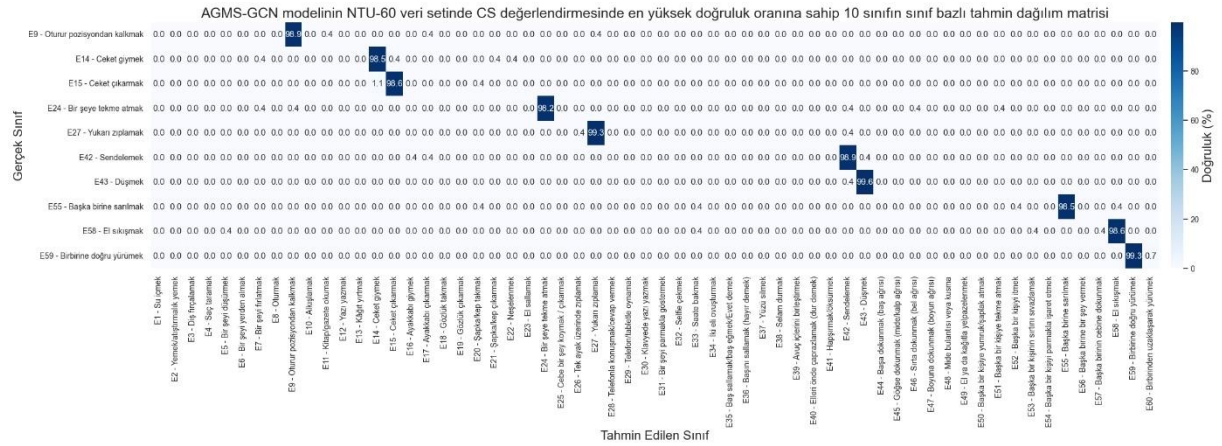
Şekil 25’de AGMS-GCN mimarisinin NTU-60 veri kümesinde CS değerlendirme protokolüne göre elde ettiği normalize edilmiş (% olarak) karışıklık matrisi sunulmuştur. Model, farklı bireyler arasında uzaysal-zamansal örüntüleri öğrenerek %93,9 genel doğruluk oranına ulaşmıştır. Bu başarı, mimarinin dikkatle yönlendirilen mekanizmasına ve çok ölçekli çizge yapısına dayanan özellik çıkarım kapasitesinden kaynaklanmaktadır.



Şekil 25. AGMS-GCN mimarisinin NTU-60 veri kümesinde CS protokolü kapsamında elde ettiği, tüm eylem sınıflarını içeren ve hücre değerleri sınıf bazında yüzdelik (%) olarak normalize edilmiş karışıklık matrisi

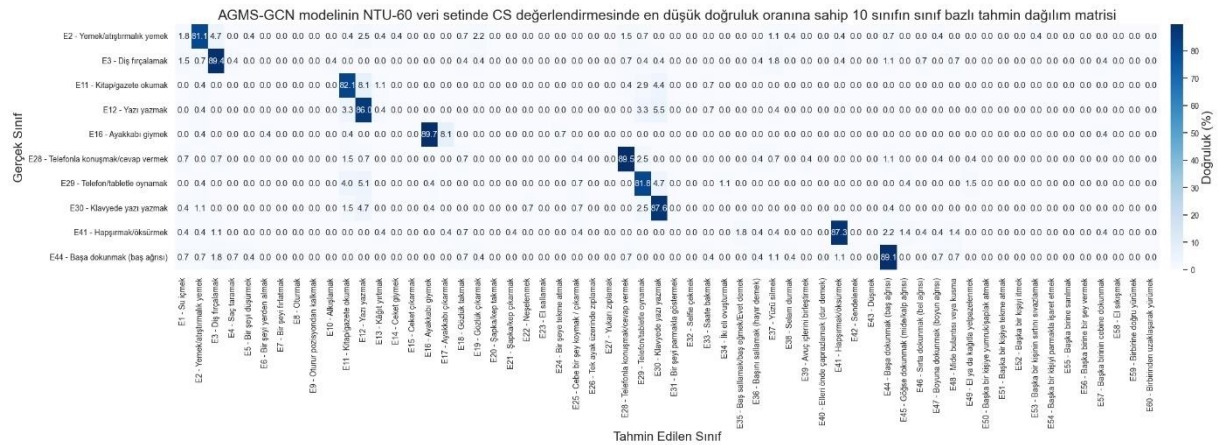
AGMS-GCN, uzaysal-zamansal çizge evrişimleri tarafından çıkarılan özellik haritalarına dikkat mekanizması uygulayarak, eylem özgü yeni bitişiklik matrisleri oluşturur. Bu sayede hem kısa hem de uzun menzilli bağıntılar yakalanır. Şekil 26’da AGMS-GCN modelinin CS değerlendirmesinde en yüksek doğruluk oranına sahip 10 sınıfın sınıf bazlı tahmin dağılım matrisi yer almaktadır. Matris incelendiğinde, “E43- Düşmek” ve “E27- Yukarı zıplamak” eylemlerinin %99’un üzerinde doğrulukla sınıflandırıldığı görülmektedir. Benzer şekilde, “E59- Birbirine doğru yürümek”, “E42- Sendelemek”, “E9- Oturur pozisyonundan kalkmak” gibi eylemler de yüksek başarı oranlarına sahiptir. Bu eylemler, bütün vücut hareketi değişiklikleri içerdiğinden dolayı, dikkat mekanizmasının uzak eklem etkileşimlerini

vurgulaması ve çok ölçekli yapının hem yerel hem de global ilişkileri modellemesi sayesinde model tarafından kolaylıkla ayırt edilebilmektedir.



Şekil 26. AGMS-GCN modelinin NTU-60 veri kümesinde, CS protokolü altında en yüksek doğruluk oranına sahip 10 sınıfın sınıf bazlı tahmin dağılım matrisi

Şekil 27’de, AGMS-GCN modelinin CS değerlendirmesinde en düşük doğruluk oranına sahip 10 sınıfın sınıf bazlı tahmin dağılım matrisi yer almaktadır. Bu sınıflar arasında “E2- Yemek/atıştırmalık yemek”, “E29- Telefon/tabletle oynamak”, “E11- Kitap/gazete okumak”, “E12- Yazı yazmak” ve “E41- Hapşırma/öksürmek” gibi eylemler yer almaktadır. Bu eylemler, düşük genlikli ve birbirine benzeyen el/kol hareketleri içermeleri nedeniyle modelin dikkat mekanizması tarafından yeterince ayrıştırılamamış ve sınıflar arası karışıklık meydana gelmiştir. Ayrıca bireyler arasında büyük varyans gösteren örüntülere sahip olmaları da modelin ayrıştırma kapasitesini zorlaştırmıştır.



Şekil 27. AGMS-GCN modelinin NTU-60 veri kümesinde, CS protokolü altında en düşük doğruluk oranına sahip 10 sınıfın sınıf bazlı tahmin dağılım matrisi

AGMS-GCN’nin dikkatle yönlendirilen çizge evrişimleri, uzak eklemler arasındaki bağlamı öğrenerek eyleme özgü ilişkilerin çıkarılmasını sağlar. Bu yapı, farklı hareket türleri arasında değişen bağlantı kalıplarını öğrenerek modele genelleme kapasitesi kazandırmaktadır. Ayrıca çok ölçekli yapısı sayesinde farklı düzeydeki ilişkilerin modellenmesi mümkün hale

gelir. Böylece AGMS-GCN, yalnızca geniş hareketleri değil, aynı zamanda karmaşık koordinasyon gerektiren eylemleri de yüksek doğrulukla sınıflandırabilmektedir. Ancak elde edilen sonuçlar, detaylı ince motor hareketlerin daha yüksek çözünürlüklü ya da bölgeye özgü dikkat odaklarıyla desteklenmesinin modelin başarısını daha da artırabileceğini göstermektedir.

Son teknoloji yöntemlerle karşılaştırmalı doğruluk analizi

Önceki bölümlerde, çok ölçekli yaklaşımın ve çoklu modalite öğrenmenin modelin performansı üzerindeki etkisini inceledik. Bu bölümde, nihai modelimizi mevcut literatürdeki diğer son teknoloji yöntemleriyle karşılaştırıyoruz. Tablo 8 ve Tablo 9’da, son teknoloji yöntemlerle NTU-60, NTU-120 ve N-UCLA veri kümeleri üzerinde elde edilen Top-1 doğruluk sonuçları karşılaştırılmaktadır. Top-1 doğruluk, modelin en yüksek olasılığa sahip tahmininin gerçek etiket ile tam olarak eşleşmesini ifade eden geleneksel doğruluk ölçütüdür. Bu tablolarda akış tipi, çalışmanın yayınladığı konferans veya dergisi, yayın yılı ve değerlendirme protokollerinde elde edilen doğruluk sonuçları verilmektedir. Tablolar, önerdiğiniz AGMS-GCN modeli ile diğer yöntemler arasında doğrudan bir karşılaştırma yapılmasına olanak tanımaktadır. Adil bir değerlendirme karşılaştırma sürecinde yalnızca GCN tabanlı yaklaşımlar dikkate alınmıştır. Kalın ve altı çizili değerler sırasıyla her veri kümesi için en iyi ve ikinci en iyi sonuçları göstermektedir. Listelenen yöntemlerin doğruluk oranları ilgili makalelerinden alınmıştır. Tablolarda bulunan akış sütünü birleştirilen modalite sayısını göstermektedir. Örneğin önerdiğimiz yöntem eklem, kemik, eklem hareket ve kemik hareket olmak üzere dört akışın (4a) skorlarını toplanmasıyla nihai sonuç elde edilmektedir.

Tablo 8. AGMS-GCN Mimarisinin NTU-60 ve NTU-120 Veri Kümeleri Üzerinde Son Teknoloji Yöntemlerle Top-1 Doğruluk (%) Analizi

Yöntemler	Akış	Yayın	Tarih	NTU-60		NTU-120	
				CS	CV	CS	CE
ST-GCN (Yan <i>et al.</i> 2018)	1a	AAAI	2018	81,5	88,3	70,7	73,2
2s-AGCN (Shi <i>et al.</i> 2019b)	2a	CVPR	2019	88,5	95,1	82,5	84,2
AGC-LSTM (Si <i>et al.</i> 2019)	2a	CVPR	2019	89,2	95,0	-	-
DGNN (Shi <i>et al.</i> 2019a)	2a	CVPR	2019	89,9	96,1	-	-
AS-GCN (Li <i>et al.</i> 2019)	1a	CVPR	2019	86,8	94,2	-	-
SGN (Zhang <i>et al.</i> 2020)	2a	CVPR	2020	89,0	94,5	79,2	81,5
STA-GCN (Shiraki <i>et al.</i> 2020)	4a	ACCV	2020	90,1	95,6	83,9	86,5
Shift-GCN (Cheng <i>et al.</i> 2020a)	4a	CVPR	2020	90,7	96,5	85,9	87,6
DC-GCN (Cheng <i>et al.</i> 2020b)	4a	ECCV	2020	90,8	96,6	86,5	88,1
DDGCN (Korban and Li 2020)	3a	ECCV	2020	91,1	97,1	-	-
MS-AAGCN (Shi <i>et al.</i> 2020)	4a	TIP	2020	90,0	96,2	-	-

Tablo 8. (devamı)

DynamicGCN (Ye <i>et al.</i> 2020)	4a	ACM	2020	91,5	96,0	87,3	88,6
MS-G3D (Liu <i>et al.</i> 2020)	2a	CVPR	2020	91,5	96,2	86,9	88,4
FGCN (Yang <i>et al.</i> 2021)	3a	TIP	2021	90,2	96,3	85,4	87,4
MST-GCN (Chen <i>et al.</i> 2021)	4a	AAAI	2021	91,5	96,6	87,5	88,8
CTR-GCN (Chen <i>et al.</i> 2021)	4a	ICCV	2021	92,4	96,8	88,9	90,6
EfficientGCN (Song <i>et al.</i> 2022)	4a	TPAMI	2022	92,1	96,1	88,7	88,9
STF (Ke <i>et al.</i> 2022)	2a	AAAI	2022	92,5	96,9	88,9	89,9
InfoGCN (Chi <i>et al.</i> 2022)	4a	CVPR	2022	92,7	96,9	89,4	90,7
InfoGCN (Chi <i>et al.</i> 2022)	6a	CVPR	2022	93,0	97,1	89,8	91,2
ML-STGNet (Zhu <i>et al.</i> 2022)	2a	TIP	2022	91,9	96,2	88,6	90,0
FR Head (Zhou <i>et al.</i> 2023)	4a	CVPR	2023	92,8	96,8	89,5	90,9
HD-GCN (Lee <i>et al.</i> 2023)	4a	ICCV	2023	93,0	97,0	89,8	91,2
HD-GCN (Lee <i>et al.</i> 2023)	6a	ICCV	2023	93,4	97,2	90,1	<u>91,6</u>
BlockGCN (Zhou <i>et al.</i> 2024)	4a	CVPR	2024	93,1	97,0	90,3	91,5
DS-GCN (Xie <i>et al.</i> 2024)	4a	AAAI	2024	93,1	<u>97,5</u>	89,2	91,1
HA-GNN (Geng <i>et al.</i> 2024)	3a	TMM	2024	93,4	97,2	89,9	91,5
SelfGCN (Wu <i>et al.</i> 2024)	4a	TIP	2024	93,1	96,6	89,4	91,0
DeGCN (Myung <i>et al.</i> 2024)	4a	TIP	2024	<u>93,6</u>	97,4	91,0	92,1
AGMS-GCN (G₂₅)	4a	KBS	2025	93,5	<u>97,5</u>	89,6	90,9
AGMS-GCN (G₆, G₁₁, G₂₅)	4a	KBS	2025	93,9	97,8	<u>90,4</u>	<u>91,6</u>

Tablo 8 incelediğinde önerilen AGMS-GCN modelinin mevcut son teknoloji yöntemlere kıyasla üstün performans sergilediğini göstermektedir. NTU-60 veri kümesinde, önerilen AGMS-GCN (G₆, G₁₁, G₂₅) modeli, CS ve CV değerlendirmelerinde sırasıyla %93,9 ve %97,8 doğruluk oranlarıyla diğer tüm yöntemleri geride bırakmıştır. Benzer şekilde güçlü bir performans NTU-120 veri kümesinde göstermiştir. Önerilen model, CS ve CE değerlendirmelerinde sırasıyla %90,4 ve %91,6 doğruluk oranlarına ulaşarak mevcut son teknoloji yöntemlerle rekabetçi sonuçlar elde etmiştir. Bu, modelimizin büyük ve karmaşık veri kümelerinde etkili bir şekilde çalıştığını göstermektedir.

N-UCLA veri kümesinde ise, AGMS-GCN modeli %97,4 doğrulukla en iyi doğruluk sonucunu elde etmiştir. Bu sonuçlar, önerilen AGMS-GCN modelinin iskelet tabanlı eylem tanımadaki eyleme özgü eklemeler arasındaki hem kısa menzilli hem de uzun menzilli uzaysal-zamansal ilişkileri daha iyi modelleyebileceğini ve böylece eylem tanıma performansını iyileştirebileceğini doğrulamaktadır. AGMS-GCN mimarisi, çok ölçekli yapıların birleştirilmesi ve dikkat tabanlı bitişiklik matrislerinin entegrasyonu sayesinde mevcut yöntemlere kıyasla daha yüksek doğruluk oranlarına ulaşmıştır. Özellikle NTU-60 ve N-UCLA

veri kümeleri üzerinde ulaşılan sonuçlar, modelin genelleme kapasitesinin yüksek olduğunu ve insan eylemlerini tanımda güçlü bir performans sergilediğini göstermektedir.

Tablo 9. AGMS-GCN Mimarisinin N-UCLA Veri Kümesi Üzerinde Son Teknoloji Yöntemlerle Top-1 Doğruluk (%) Analizi

Yöntemler	Akış	Yayın	Tarih	CV
AGC-LSTM (Si <i>et al.</i> 2019)	2a	CVPR	2019	93,3
SGN (Zhang <i>et al.</i> 2020)	2a	CVPR	2020	92,5
Shift-GCN (Cheng <i>et al.</i> 2020a)	4a	CVPR	2020	94,6
DC-GCN+ADG (Cheng <i>et al.</i> 2020b)	4a	ECCV	2020	95,3
CTR-GCN (Chen <i>et al.</i> 2021)	4a	ICCV	2021	96,5
InfoGCN (Chi <i>et al.</i> 2022)	4a	CVPR	2022	96,6
InfoGCN (Chi <i>et al.</i> 2022)	6a	CVPR	2022	97,0
FR Head (Zhou <i>et al.</i> 2023)	4a	CVPR	2023	96,8
HD-GCN (Lee <i>et al.</i> 2023)	4a	ICCV	2023	96,9
HD-GCN (Lee <i>et al.</i> 2023)	6a	ICCV	2023	<u>97,2</u>
BlockGCN (Zhou <i>et al.</i> 2024)	4a	CVPR	2024	96,9
SelfGCN (Wu <i>et al.</i> 2024)	4a	TIP	2024	96,8
DeGCN (Myung <i>et al.</i> 2024)	4a	TIP	2024	<u>97,2</u>
AGMS-GCN (G_{25})	4a	KBS	2025	97,1
AGMS-GCN (G_6, G_{11}, G_{25})	4a	KBS	2025	97,4

FCSA-GCN Mimarisinin Deney Sonuçları

Bu bölümde, önerilen FCSA-GCN mimarisinin farklı bileşenlerinin etkisini değerlendirmek amacıyla ayrıntılı bir ablasyon çalışması gerçekleştirilmiştir. İlk olarak, inceden-kabaya yaklaşımının etkinliği, farklı ölçeklerdeki çizge yapılarının birleştirilmesinin model performansına olan katkısı deneysel olarak incelenmiştir. Daha sonra, toplam kayıp değerinin optimizasyonun doğruluk üzerine etkisi incelenmiştir. Ardından, zamansal öz-dikkat mekanizmasının zamansal özelliklerin çıkarımındaki rolü ve mimarinin doğruluğuna olan etkisi analiz edilmiştir. Son olarak, önerilen yöntemin performansı, mevcut son teknoloji yöntemlerle karşılaştırılarak genel başarıyı değerlendirilmiştir. Bu analizler ile mimarinin temel yeniliklerini ve bileşenlerinin uygulamadaki önemini vurgulamayı amaçlanmaktadır.

İnceden-kabaya yaklaşım ile farklı ölçeklerde doğruluk analizi

Modelin inceden-kabaya özellik çıkarma yaklaşımının NTU-60 veri kümesindeki performansı, farklı ölçeklerden ve bunların kombinasyonlarından türetilen çizge yapıları kullanılarak Tablo 10 ve Tablo 11’de sunulmuştur. Bu sonuçlar önerilen inceden-kabaya yaklaşımının etkinliğini açıkça göstermektedir. CS ve CV değerlendirme protokollerinde en iyi

izge lek yapılarının G_6 , G_{25} olduėu grlmektedir. G_{11} izge lek yapısı doėruluk baėlamında diėer izge yapılarının gerisinde kalmaktadır. Ancak lekler birleřtirildiėinde, tm modalitelerde nemli bir doėruluk iyileřtirmesi elde edilmektedir. rneėin eklem modalitesi iin, inceden-kabaya řeklinde birleřtirilmiř lek yapısı CS deėerlendirme protokolnde doėruluėu %88,36'ya CV deėerlendirme protokolnde ise doėruluėu %94,14'e ykseltmektedir.

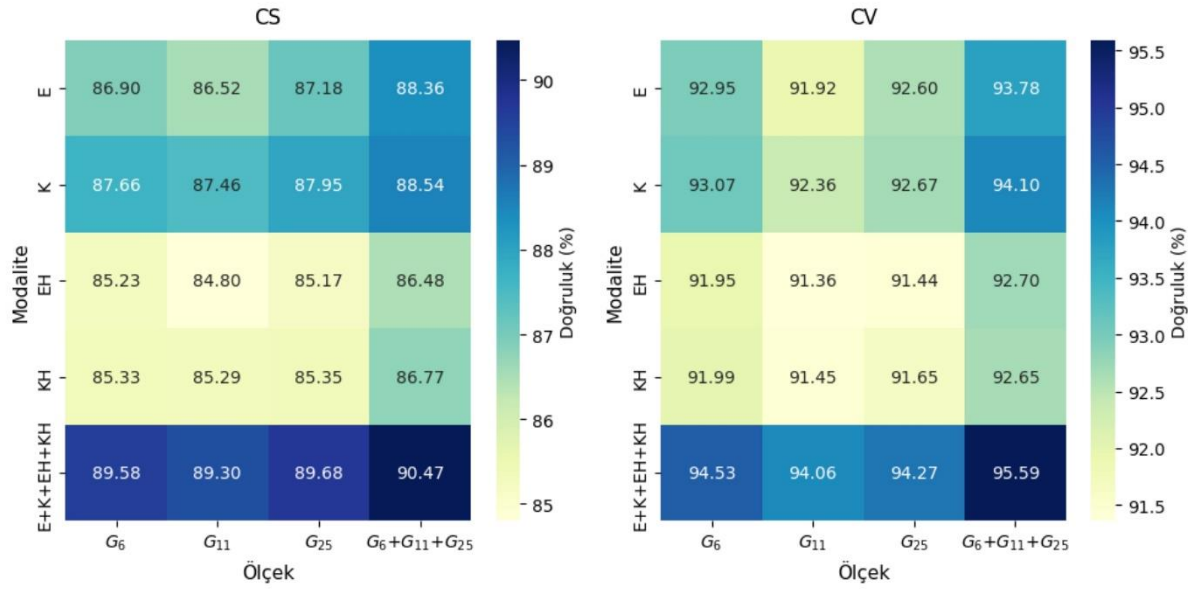
Tablo 10. FCSA-GCN Mimarisinin NTU-60 Veri Kmesinde CS Protokolne Gre, Farklı lek ve Modalitelerde İnceden-Kabaya Yaklařımıyla Elde Ettiėi Doėruluk Oranları (%)

Modalite	G_6	G_{11}	G_{25}	$G_6 + G_{11} + G_{25}$
Eklem	86,90	86,52	87,18	88,36
Kemik	87,66	87,46	87,95	88,54
Eklem Hareket	85,23	84,80	85,17	86,48
Kemik Hareket	85,33	85,29	85,35	86,77
E + K + EH + KH	89,58	89,30	89,68	90,47

Tablo 11. FCSA-GCN Mimarisinin NTU-60 Veri Kmesinde CV Protokolne Gre, Farklı lek ve Modalitelerde İnceden-Kabaya Yaklařımıyla Elde Ettiėi Doėruluk Oranları (%)

Modalite	G_6	G_{11}	G_{25}	$G_6 + G_{11} + G_{25}$
Eklem	92,95	91,92	92,60	93,78
Kemik	93,07	92,36	92,67	94,10
Eklem Hareket	91,95	91,36	91,44	92,70
Kemik Hareket	91,99	91,45	91,65	92,65
E + K + EH + KH	94,53	94,06	94,27	95,59

Hareket trevi modalitelerde bireysel leklerin doėruluk oranları daha dřk olsa da birleřik lekler kullanıldıėında bu modalitelerde de nemli iyileřtirmeler gzlenmektedir. Tm modalitelerin topluluk kombinasyonu en yksek doėruluk oranlarına ulařmaktadır. G_{25} leėinde %89,68 olan CS doėruluėu, birleřik leklerle %90,47'ye ykselmektedir. Benzer řekilde, CV doėruluėu %94,27'den %95,59'a ykselmektedir. Sonu olarak, nerilen inceden-kabaya yaklařım, farklı leklerde uzaysal-zamansal iliřkilerin entegre edilmesinin eylem tanıma performansı zerindeki olumlu etkisini aıka grlmektedir. zellikle, birleřik leklerin bireysel leklere kıyasla daha gl bir temsil saėladıėı ve modelin genel performansını artırdıėı gzlenmektedir. Tablo 10 ve Tablo 11'de sunulan deneysel sonular, daha net bir temsil iin bir ısı haritaları kullanılarak řekil 28'de grselleřtirilmiřtir. Bu ablasyon alıřmasında, nihai modelde bulunan toplam kayıp deėerinin DWA ile optimizasyonu ve zamansal z-dikkat mekanizması bileřenleri hari tutulmuřtur. Bu bileřenlerin performans zerindeki etkisi sonraki deėerlendirmelerde ayrıntılı olarak analiz edilmiřtir.



Şekil 28. Modalite ve ölçeğe göre CS/CV değerlendirme protokollerinde doğrulukları

Toplam kayıp değerinin optimizasyonun doğruluk üzerine etkisi

Tablo 12 ve Tablo 13'te, NTU-60 veri kümesi üzerinde inceden-kabaya ve DWA yaklaşımlarıyla farklı ölçek ve modalitelerde, CS ve CV protokollerine göre elde edilen doğruluk oranları (%) sunulmaktadır.

Tablo 12. FCSEA-GCN Mimarısının NTU-60 Veri Kümesinde İnceden-Kabaya ve DWA Yaklaşımlarıyla, Farklı Ölçek ve Modalitelerde CS Protokolüne Göre Elde Ettiği Doğruluk Oranları (%)

Modalite	G ₆	G ₁₁	G ₂₅	G ₆ + G ₁₁ + G ₂₅
Eklem	87,33	87,39	87,67	88,68
Kemik	88,13	88,28	88,49	88,93
Eklem Hareket	85,61	85,59	85,66	86,84
Kemik Hareket	85,82	86,07	85,87	87,18
E + K + EH + KH	90,10	90,23	90,42	91,08

Tablo 13. FCSEA-GCN Mimarısının NTU-60 Veri Kümesinde İnceden-Kabaya ve DWA Yaklaşımlarıyla, Farklı Ölçek ve Modalitelerde CV Protokolüne Göre Elde Ettiği Doğruluk Oranları (%)

Modalite	G ₆	G ₁₁	G ₂₅	G ₆ + G ₁₁ + G ₂₅
Eklem	93,34	92,90	93,14	94,14
Kemik	93,49	93,14	93,28	94,44
Eklem Hareket	92,31	92,23	92,01	93,03
Kemik Hareket	92,39	92,27	92,18	93,07
E + K + EH + KH	94,97	94,73	94,82	96,02

Tabloları dikkate alındığında, DWA yönteminin doğruluk üzerinde olumlu bir etkisi olduğu açıkça görülmektedir. Özellikle farklı modalitelerin ve çizge yapı ölçeklerinin bir arada kullanıldığı durumda, DWA'nın sağladığı dinamik ağırlıklandırma yaklaşımı sayesinde modelin genel başarımı artmıştır. Mimari yapıya DWA'nın eklenmesiyle CS değerlendirme protokolünde doğruluk %90,47'den %91,08 CV değerlendirme protokolünde ise %95,59'dan %96,02 artmıştır. Bu artış, DWA'nın modelin mimarisinde herhangi bir değişikliğe yol açmadan yalnızca eğitim sürecinde kayıp fonksiyonlarının ağırlıklarını dinamik olarak ayarlamasıyla elde edilmektedir.

DWA yalnızca çoklu kayıp fonksiyonlarının ağırlıklarını dinamik olarak günceller. Bu ağırlıklar, modelin öğrenme sürecini yönlendiren hesaplamalar sırasında kullanılır. Ancak bu ağırlıklar modelin kendisinde tanımlı parametreler değildir; yani, modelin toplam eğitilebilir parametreleri ve mimarisi sabit kalır. Aynı şekilde, modelin hesaplama karmaşıklığı açısından önemli bir ölçüt olan FLOPs değeri de DWA'dan etkilenmez. FLOPs hesaplaması, yalnızca modelin katman sayısı, tipleri, filtre sayısı, çekirdek boyutu, giriş/çıkış boyutları gibi yapısal özellikler üzerinden hesaplanır. DWA ise yalnızca eğitim sürecinde, hesaplanan kayıpların önem düzeyine göre dinamik ağırlıklandırma yapar. Bu nedenle, FLOPs değeri değişmeden kalırken, modelin eğitimi sırasında yapılan bu stratejik ağırlıklandırma ile genel doğruluk iyileştirilmektedir. DWA'nın sağladığı başlıca katkı, GFE_{25} , GFE_{11} , GFE_6 özellik çıkarıcılarının kayıp fonksiyonları arasındaki dengesizlikleri gidermesidir. Her bir çıkışın öğrenme zorluğu farklı olabilir; DWA, bu farklılıkları her eğitim döngüsü sonunda yeniden değerlendirerek ağırlıkları güncelleyerek gradyan çakışmalarının azaltılmasına ve daha verimli bir optimizasyon sürecine olanak tanır. Eğitim süreci boyunca sabit ağırlıklar yerine adaptif ağırlıkların kullanılması, modelin daha iyi genelleme yapmasını sağlar ve bu da test kümesindeki doğruluğun artmasına katkı sunar. DWA eğitim süresini hafifçe uzatır fakat bu durum modelin yapısından değil, eğitim sürecindeki ek işlemlerden kaynaklanır. Eğitim sırasındaki bu ekstra hesaplama yükü ihmal edilebilir düzeydedir. Sonuç olarak, DWA'nın eğitim dinamiklerine odaklanan yaklaşımı, doğrudan model mimarisini değiştirmeden, yalnızca eğitim sürecine yönelik ince ayarlarla doğruluk değerlerinde anlamlı iyileştirme sağlamaktadır.

Zamansal öz-dikkat mekanizmasının doğruluk üzerine etkisi

Tablo 14 ve Tablo 15’te NTU-60 veri kümesi üzerinde farklı modalitelerde zamansal öz-dikkat mekanizması kullanılmadan ve kullanılarak elde edilen doğruluk oranları karşılaştırmaktadır.



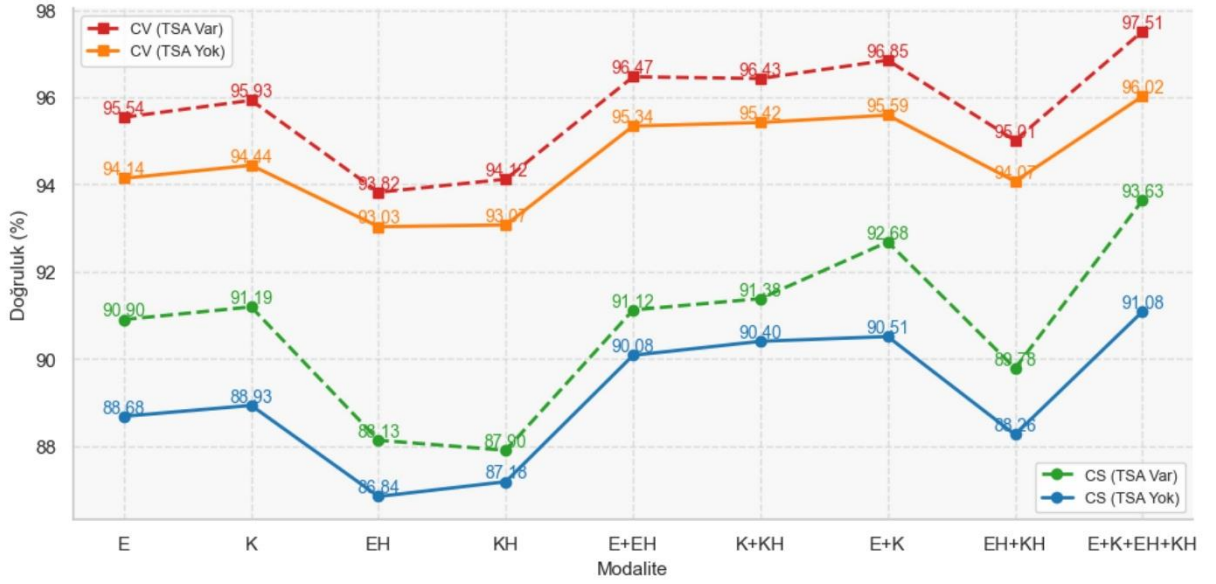
Tablo 14. FCSA-GCN Mimarisinin NTU-60 Veri Kümesinde Farklı Modaliteler İçin TSA Olmadan Doğruluk (%) Karşılaştırması

Modalite	CS	CV
Eklem	88,68	94,14
Kemik	88,93	94,44
Eklem Hareket	86,84	93,03
Kemik Hareket	87,18	93,07
Eklem + Eklem Hareket	90,08	95,34
Kemik + Kemik Hareket	90,40	95,42
Eklem + Kemik	90,51	95,59
Eklem Hareket + Kemik Hareket	88,26	94,07
E + K + EH + KH	91,08	96,02

Tablo 15. FCSA-GCN Mimarisinin NTU-60 Veri Kümesinde Farklı Modaliteler İçin TSA ile Doğruluk (%) Karşılaştırması

Modalite	CS	CV
Eklem	90,90	95,54
Kemik	91,19	95,93
Eklem Hareket	88,13	93,82
Kemik Hareket	87,90	94,12
Eklem + Eklem Hareket	91,12	96,47
Kemik + Kemik Hareket	91,38	96,43
Eklem + Kemik	92,68	96,85
Eklem Hareket + Kemik Hareket	89,78	95,01
E + K + EH + KH	93,63	97,51

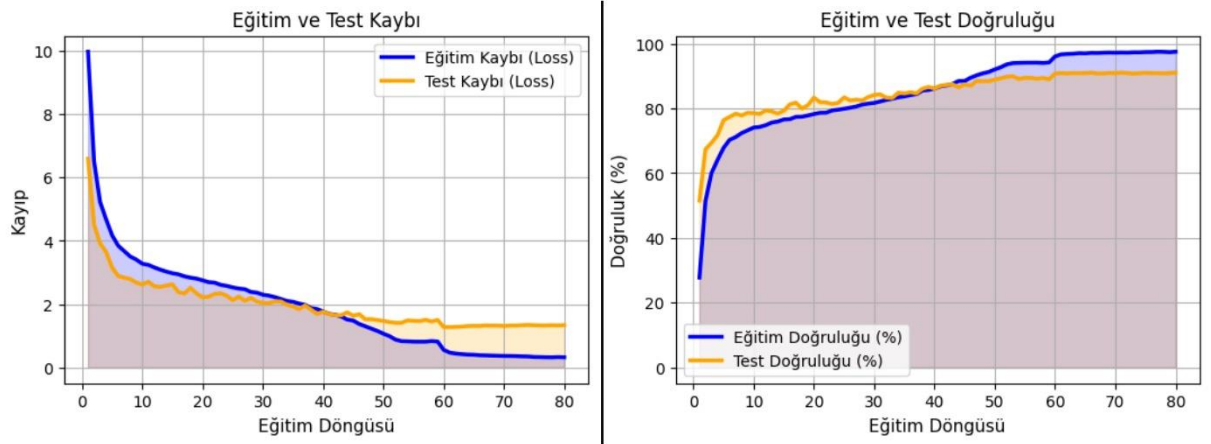
Tablolar incelendiğinde, TSA mekanizmasının uygulanması ile model performansında tüm modaliteler ve kombinasyonlar için dikkate değer bir iyileşme sağlandığı görülmektedir. Özellikle, TSA'nın etkisi tüm modalitelerin birleşiminde: CS doğruluğu %91,08'den %93,63'e ve CV doğruluğu %96,02'den %97,51'e yükselmektedir. TSA'nın katkısı bireysel modalitelerde de belirgindir. Eklem modalitesi için TSA, CS doğruluğunu %88,68'den %90,90'a ve CV doğruluğunu %94,14'ten %95,54'e çıkarmaktadır. Benzer şekilde, kemik modalitesi için TSA CS doğruluğunu %88,93'ten %91,19'a ve CV doğruluğunu %94,44'ten %95,93'e çıkarmaktadır. Hareketten türetilen modaliteler için de benzer doğruluk iyileşmeleri gözlenmektedir. Bu sonuçlar TSA'nın zamansal özelliklerin daha etkili bir şekilde yakalanmasını sağladığını ve modelin performansını optimize ettiğini ortaya koymaktadır. Şekil 29'da TSA'nın dahil edilmesiyle elde edilen doğruluk iyileştirmelerinin karşılaştırmalı bir görselleştirmesi verilmiştir.



Şekil 29. TSA ile/olmadan farklı modalitelerde doğruluk.

Öğrenme sürecinin görsel analizi

Bu bölümde FCSA-GCN mimarisinin nihai modeline ait eğitim grafikleri incelenmektedir. Şekil 30’da NTU-60 veri kümesinde CS değerlendirme protokolü kapsamında eklem modalitesinde elde edilen eğitim süreci grafikleri sunulmaktadır.

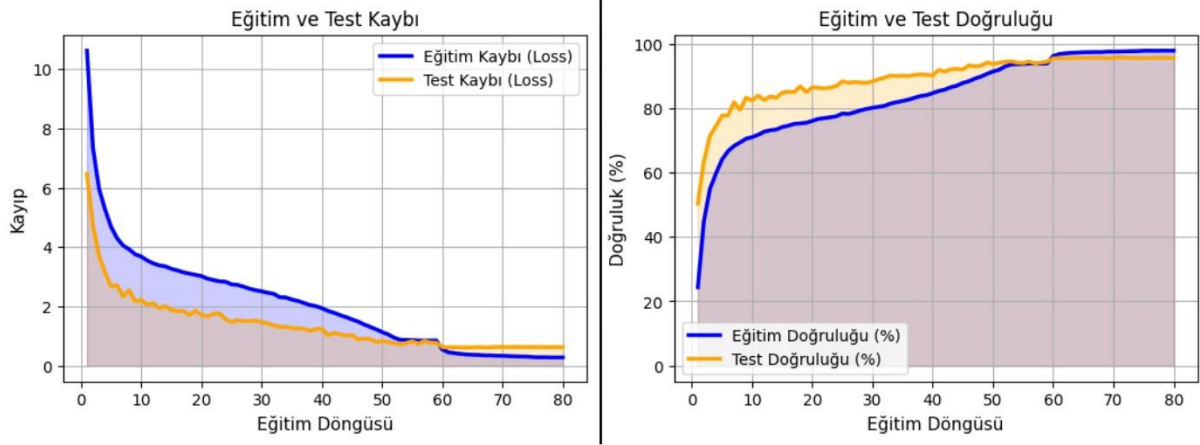


Şekil 30. FCSA-GCN mimarisinin NTU-60 veri kümesi üzerinde CS protokolü kapsamında eklem modalitesine ait eğitim grafikleri

CS senaryosundaki eğitim süreci incelendiğinde, eğitim kaybı eğrisinin sürekli olarak azaldığı, eğitim doğruluğunun ise kararlı bir şekilde arttığı gözlemlenmektedir. Ancak, test kaybında iyileşmenin durduğu ve test doğruluğunun plato yaptığı görülmüştür. Eğitim sürecinde erken durdurma sabır değeri 7 olarak belirlenmiş olup, 61. eğitim döngüsü sonrasında 7 ardışık eğitim döngüsü (62-68) boyunca test performansında anlamlı bir iyileşme sağlanamamıştır. Bu nedenle eğitim süreci 68. eğitim döngüsünde erken durdurma uygulanarak sonlandırılmıştır. Grafikte her ne kadar 80 eğitim döngüsü yer alsada nihai model 61. eğitim döngüsünde kaydedilmiştir. Ancak, eğitim süreci başlangıç parametrelerinde 80 eğitim

döngüsü olarak tanımlandığı için, modelin performans eğiliminin daha ayrıntılı gözlemlenebilmesi amacıyla eğitim tamamlanana kadar sürdürülmüştür.

Şekil 31’de NTU-60 veri kümesinde CV değerlendirme protokolü kapsamında eklem modalitesinde elde edilen eğitim süreci grafikleri sunulmuştur.



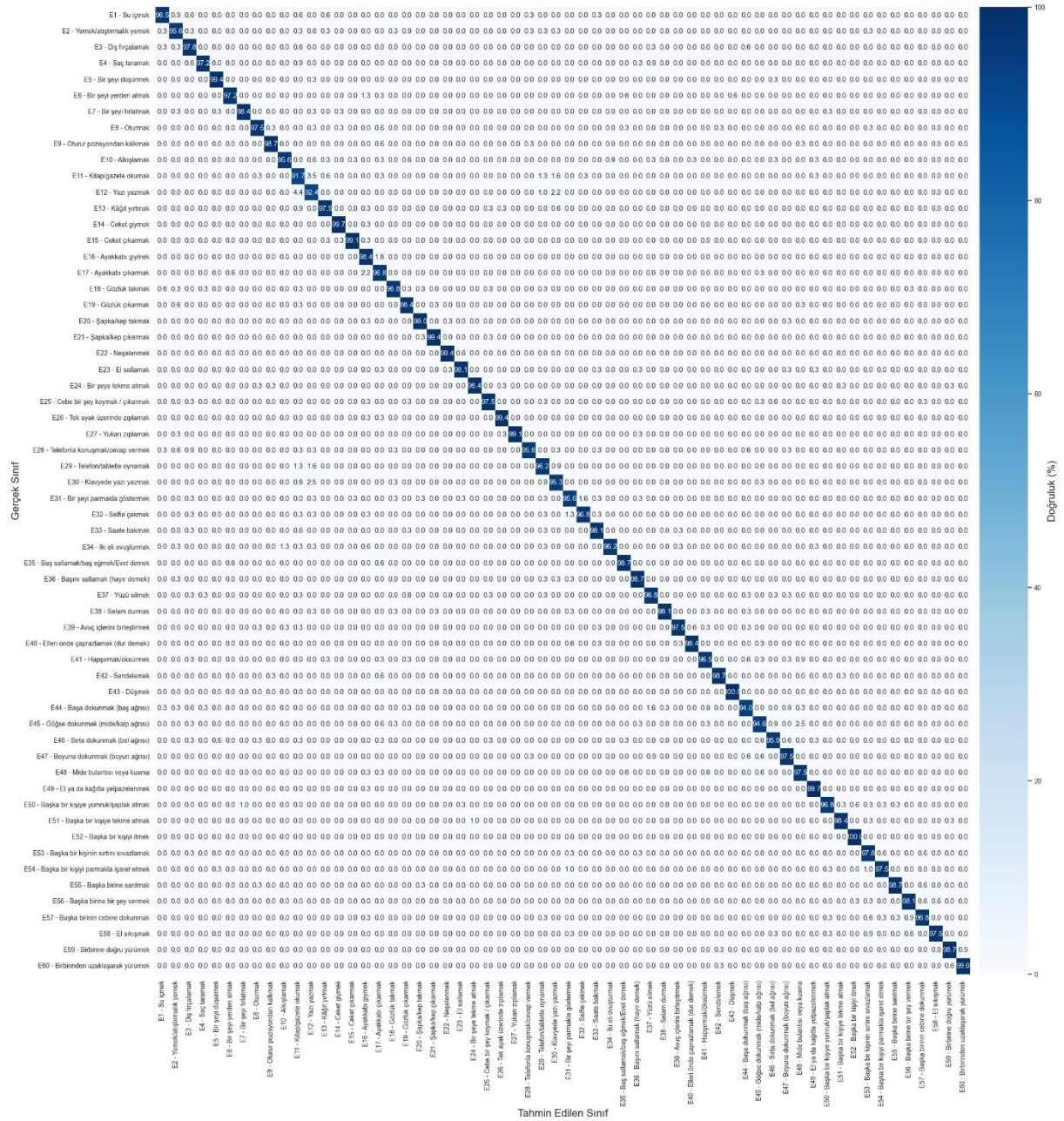
Şekil 31. FCSA-GCN mimarisinin NTU-60 veri kümesi üzerinde CV protokolü kapsamında eklem modalitesine ait eğitim grafikleri

Şekil 31’deki CV senaryosunun eğitim süreci incelendiğinde, eğitim kaybı eğrisinin düzenli olarak azaldığı ve eğitim doğruluğunun istikrarlı bir şekilde arttığı gözlemlenmektedir. Benzer şekilde, test doğruluğu da başlangıçta yükselmiş ancak 64. eğitim döngüsünden itibaren plato yaparak sabitlenmiştir. Bu eğitim döngüsü sonrasında test kaybında anlamlı bir iyileşme gözlenmemiş ve doğruluk değerleri stabil hale gelmiştir. Eğitim sürecinde erken durdurma sabır değeri 7 olarak belirlenmiş olup, 64. eğitim döngüsü sonrasında 7 ardışık eğitim döngüsü (65-71) boyunca test performansında ilerleme kaydedilememiştir. Bu nedenle eğitim süreci 71. eğitim döngüsünde erken durdurma uygulanarak sonlandırılmıştır. Bu durum, modelin aşırı öğrenmeye geçmeden en yüksek genelleme performansına ulaştığı noktada eğitimin başarıyla tamamlandığını göstermektedir. Grafikte 80 eğitim döngüsü yer almakta olup, nihai model 64. eğitim döngüsünde kaydedilmiştir.

Mimarinin sınıf bazlı değerlendirmesi

Şekil 32’de, FCSA-GCN mimarisinin NTU-60 veri kümesinde CV değerlendirme protokolüne göre elde ettiği normalize edilmiş (% olarak) karışıklık matrisi sunulmuştur. Model, bakış açısı değişikliklerine karşı %97,51 genel doğruluk oranına ulaşmıştır. Özellikle vücut postüründe belirgin değişiklikler içeren ya da tüm vücutla gerçekleştirilen hareketlerde modelin neredeyse hatasız çalıştığı gözlemlenmiştir. Öte yandan, yazı yazmak, kitap okumak, başa veya sırtta dokunmak gibi ince motor beceriler gerektiren ve görsel olarak benzerlik gösteren eylemler arasında bazı karışıklıklar oluşmuştur. Bu durum, modelin detaylı el/kol

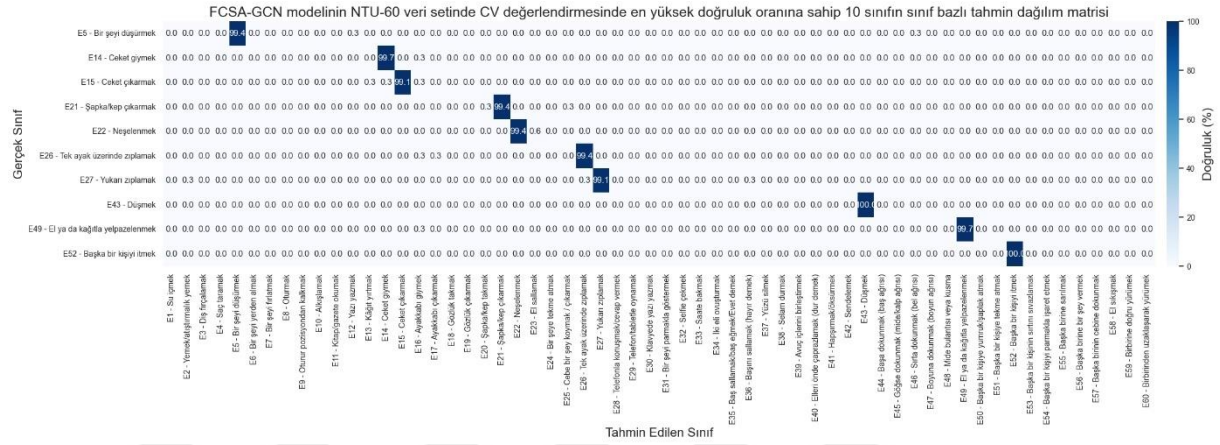
hareketlerini ayırt etme konusunda görece daha fazla zorlandığını ve ince motor hareketlere yönelik temsil yeteneğinin geliştirilebileceğini göstermektedir.



Şekil 32. FCSCA-GCN mimarisinin NTU-60 veri kümesinde CV protokolü kapsamında elde ettiği, tüm eylem sınıflarını içeren ve hücre değerleri sınıf bazında yüzdelik (%) olarak normalize edilmiş karışıklık matrisi

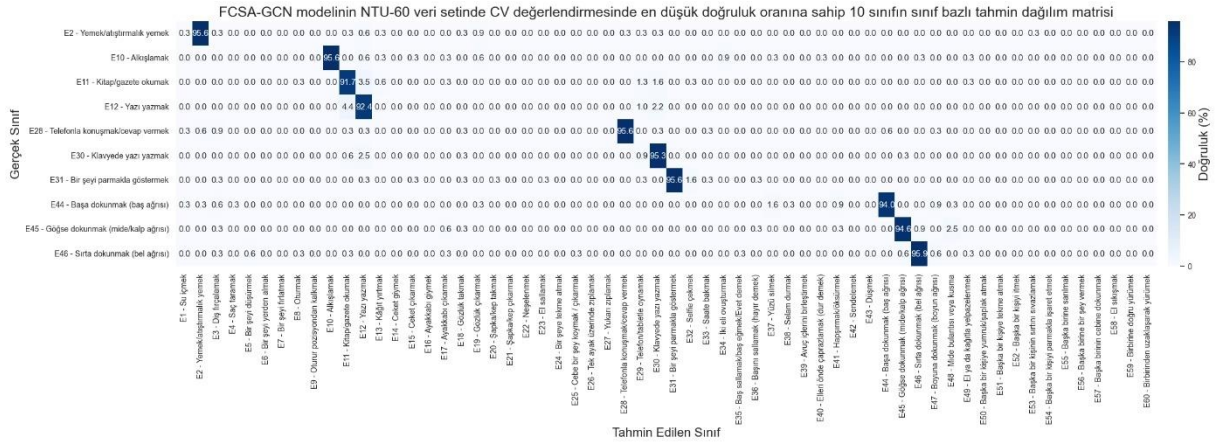
Şekil 33’de FCSCA-GCN modelinin NTU-60 veri kümesinde CV değerlendirmesinde en yüksek doğruluk oranına sahip 10 sınıfın sınıf bazlı tahmin dağılım matrisi verilmiştir. Matris incelediğinde “E52- Başka bir kişiyi itmek” ve “E43- Düşmek” eylemleri %100 doğrulukla sınıflandırılmıştır. Bu eylemler, geniş vücut hareketleri içerdiği için model tarafından kolayca ayırt edilebilmektedir. %99,4–99,7 doğruluğa sahip sınıflar genellikle koordineli eklem hareketleri içeren, büyük ölçekli fiziksel eylemlerdir. Örneğin “E14- Ceket giymek”, “E26- Tek

ayak üzerinde zıplamak” gibi eylemler birden fazla eklemen zamanla gelişen koordineli hareketini gerektirir. Modelin çok ölçekli yapısı ve zamansal öz-dikkat mekanizması bu tür eylemlere ait örüntüleri başarıyla modelleyebilmektedir. Bu durum önerilen mimarinin uzaysal-zamansal ilişkileri güçlü şekilde modellediğini ve öz-dikkat mekanizması zaman içinde gelişen eylemler için yüksek doğruluk oranlarının elde edilmesine katkı sağladığını göstermektedir.



Şekil 33. FCSA-GCN modelinin NTU-60 veri kümesinde, CV protokolü altında en yüksek doğruluk oranına sahip 10 sınıfın sınıf bazlı tahmin dağılım matrisi

Şekil 34’de FCSA-GCN modelinin NTU-60 veri kümesinde CV değerlendirmesinde en düşük doğruluk oranına sahip 10 sınıfın sınıf bazlı tahmin dağılım matrisi verilmiştir. Matristeki sınıfların ortak özelliği, çoğunun ince motor hareketleri içermesi ve/veya benzer görsel desenlere sahip olmasıdır. Örneğin, “E11- Kitap/gazete okumak”, “E12- Yazı yazmak” ve “E30- Klavyede yazı yazmak” sınıfları benzer el/kol hareketleriyle gerçekleştirilen, birbirine görsel olarak çok yakın eylemlerdir. “E44- Başa dokunmak (baş ağrısı)”, “E45- Göğse dokunmak (mide/kalp ağrısı)” ve “E46- Sırta dokunmak (bel ağrısı)” gibi sınıflar da modelin dikkatini çekmesi zor olan, ince temas hareketlerini içerir. Bu tarz hareketlerin tanınmasında, özellikle vücut segmentlerinin detaylı ayrıştırılması gerekir. Bu da genellikle daha yüksek çözünürlüklü uzaysal ayrım yeteneği gerektirir. Ayrıca, “E28- Telefonla konuşmak/cevap vermek” ve “E2- Yemek/atıştırılmalık yemek” gibi eylemler, farklı bireyler arasında büyük varyans gösterebildiğinden dolayı tutarlı bir örüntü yakalanması zor olabilmektedir.



Şekil 34. FCSA-GCN modelinin NTU-60 veri kümesinde, CV protokolü altında en düşük doğruluk oranına sahip 10 sınıfın sınıf bazlı tahmin dağılım matrisi

Son teknoloji yöntemlerle karşılaştırmalı doğruluk analiz

Önceki bölümlerde, inceden-kabaya özellik çıkarma yaklaşımın ve zamansal öz-dikkat mekanizmasının modelin performansı üzerindeki etkisini incelenmiştir. Bu bölümde, nihai modelimizi mevcut literatürdeki diğer son teknoloji yöntemleriyle karşılaştırıyoruz.

Tablo 16 ve Tablo 17’de, son teknoloji yöntemlerle NTU-60, NTU-120 ve N-UCLA veri kümeleri üzerinde elde edilen Top-1 doğruluk sonuçları karşılaştırılmaktadır. Bu tablolarda akış tipi, çalışmanın yayınladığı konferans veya dergisi, yayın yılı ve değerlendirme protokollerinde elde edilen doğruluk sonuçları verilmektedir. Tablolar, önerdiğiniz FC-SA-GCN modeli ile diğer yöntemler arasında doğrudan bir karşılaştırma yapılmasına olanak tanımaktadır. Adil bir değerlendirme karşılaştırma sürecinde yalnızca GCN tabanlı yaklaşımlar dikkate alınmıştır. Kalın ve altı çizili değerler sırasıyla her veri kümesi için en iyi ve ikinci en iyi sonuçları göstermektedir. Tablolarda listelenen yöntemlerin doğruluk oranları ilgili makalelerinden alınmıştır.



Tablo 16’da görüldüğü üzere, FCSA-GCN mimarisi CS değerlendirmesinde %93,6 CV değerlendirmesinde ise %97,5 doğruluk oranlarına ulaşarak tüm diğer yöntemlerden üstün performans sergilemiştir. Özellikle, FCSA-GCN, ST-GCN gibi geleneksel yöntemlere (CS: %81,5, CV: %88,3) ve MST-GCN gibi ileri seviye çok ölçekli yöntemlere (CS: %91,5, CV: %96,6) kıyasla belirgin bir iyileşme sağlamıştır.



Tablo 16. FCSA-GCN Mimarisinin NTU-60 ve NTU-120 Veri Kümeleri Üzerinde Son Teknoloji Yöntemlerle Top-1 Doğruluk (%) Analizi

Yöntemler	Akış	Yayın	Tarih	NTU-60		NTU-120	
				CS	CV	CS	CE
ST-GCN (Yan <i>et al.</i> 2018)	1a	AAAI	2018	81,5	88,3	70,7	73,2
2s-AGCN (Shi <i>et al.</i> 2019b)	2a	CVPR	2019	88,5	95,1	82,5	84,2
AGC-LSTM (Si <i>et al.</i> 2019)	2a	CVPR	2019	89,2	95,0	-	-
DGNN (Shi <i>et al.</i> 2019a)	2a	CVPR	2019	89,9	96,1	-	-
AS-GCN (Li <i>et al.</i> 2019)	1a	CVPR	2019	86,8	94,2	-	-
SGN (Zhang <i>et al.</i> 2020)	2a	CVPR	2020	89,0	94,5	79,2	81,5
STA-GCN (Shiraki <i>et al.</i> 2020)	6a	ACCV	2020	90,1	95,6	83,9	86,5
Shift-GCN (Cheng <i>et al.</i> 2020a)	4a	CVPR	2020	90,7	96,5	85,9	87,6
DC-GCN (Cheng <i>et al.</i> 2020b)	4a	ECCV	2020	90,8	96,6	86,5	88,1
DDGCN (Korban and Li 2020)	3a	ECCV	2020	91,1	97,1	-	-
MS-AAGCN (Shi <i>et al.</i> 2020)	4a	TIP	2020	90,0	96,2	-	-
DynamicGCN (Ye <i>et al.</i> 2020)	4a	ACM	2020	91,5	96,0	87,3	88,6
MS-G3D (Liu <i>et al.</i> 2020)	2a	CVPR	2020	91,5	96,2	86,9	88,4
FGCN (Yang <i>et al.</i> 2021)	3a	TIP	2021	90,2	96,3	85,4	87,4
MST-GCN (Chen <i>et al.</i> 2021)	4a	AAAI	2021	91,5	96,6	87,5	88,8
CTR-GCN (Chen <i>et al.</i> 2021)	4a	ICCV	2021	92,4	96,8	88,9	90,6
EfficientGCN (Song <i>et al.</i> 2022)	4a	TPAMI	2022	92,1	96,1	88,7	88,9
STF (Ke <i>et al.</i> 2022)	2a	AAAI	2022	92,5	96,9	88,9	89,9
InfoGCN (Chi <i>et al.</i> 2022)	4a	CVPR	2022	92,7	96,9	89,4	90,7
InfoGCN (Chi <i>et al.</i> 2022)	6a	CVPR	2022	93,0	97,1	89,8	91,2
ML-STGNet (Zhu <i>et al.</i> 2022)	2a	TIP	2022	91,9	96,2	88,6	90,0
FR Head (Zhou <i>et al.</i> 2023)	4a	CVPR	2023	92,8	96,8	89,5	90,9
HD-GCN (Lee <i>et al.</i> 2023)	4a	ICCV	2023	93,0	97,0	89,8	91,2
HD-GCN (Lee <i>et al.</i> 2023)	6a	ICCV	2023	<u>93,4</u>	97,2	90,1	<u>91,6</u>
BlockGCN (Zhou <i>et al.</i> 2024)	4a	CVPR	2024	93,1	97,0	90,3	91,5
DS-GCN (Xie <i>et al.</i> 2024)	4a	AAAI	2024	93,1	97,5	89,2	91,1
HA-GNN (Geng <i>et al.</i> 2024)	3a	TMM	2024	<u>93,4</u>	97,2	89,9	91,5
SelfGCN (Wu <i>et al.</i> 2024)	4a	TIP	2024	93,1	96,6	89,4	91,0
DeGCN (Myung <i>et al.</i> 2024)	4a	TIP	2024	93,6	<u>97,4</u>	91,0	92,1
FCSA-GCN	4a	-	2025	93,6	97,5	<u>90,5</u>	91,3

Tablo 16’da NTU-120 veri kümesine ait karşılaştırmalı sonuçlarda sunulmaktadır. FCSA-GCN, CS değerlendirmesinde %90,5 CE değerlendirmesinde ise %91,3 doğruluk oranlarına ulaşmıştır. Önerilen yöntem, Shift-GCN (CS: %85,9; CE: %87,6) ve CTR-GCN (CS: %88,9; CE: %90,6) gibi temel yöntemlerden daha iyi sonuçlar elde etmiş ve HD-GCN (CS: %90,1; CE: %91,6) ve DeGCN (CS: %91,0; CE: %92,1) gibi son teknoloji yöntemlerle rekabet edebilir bir performans göstermiştir. Elde edilen sonuçlar, FCSA-GCN’nin, uzaysal-zamansal özellik çıkarımını geliştirmek için hem inceden-kabaya eklem gruplamalarını hem de zamansal öz-dikkat mekanizmasını uzaysal-zamansal özellik çıkarımını geliştirmek için etkili bir şekilde kullandığını ortaya koymaktadır. Ayrıca, hem NTU-60 hem de NTU-120 gibi büyük çaplı veri kümelerinde tutarlı performans artışları, önerilen yöntemin karmaşık veri kümelerine uyarlanabilirliğini ve ölçeklenebilirliğini göstermektedir. Tablo 17’de görüldüğü üzere FCSA-GCN mimarisi N-UCLA veri kümesinde %97,2 doğrulukla en iyi sonucu elde etmiştir. Bu sonuç, modelimizin farklı veri kümelerinde genel bir performans sağladığını göstermektedir.

Tablo 17. FCSA-GCN Mimarisinin N-UCLA Veri Kümesi Üzerinde Son Teknoloji Yöntemlerle Top-1 Doğruluk (%) Analizi

Yöntemler	Akış	Yayın	Tarih	CV
AGC-LSTM (Si <i>et al.</i> 2019)	2a	CVPR	2019	93,3
SGN (Zhang <i>et al.</i> 2020)	2a	CVPR	2020	92,5
Shift-GCN (Cheng <i>et al.</i> 2020a)	4a	CVPR	2020	94,6
DC-GCN+ADG (Cheng <i>et al.</i> 2020b)	4a	ECCV	2020	95,3
CTR-GCN (Chen <i>et al.</i> 2021)	4a	ICCV	2021	96,5
InfoGCN (Chi <i>et al.</i> 2022)	4a	CVPR	2022	96,6
InfoGCN (Chi <i>et al.</i> 2022)	6a	CVPR	2022	<u>97,0</u>
FR Head (Zhou <i>et al.</i> 2023)	4a	CVPR	2023	96,8
HD-GCN (Lee <i>et al.</i> 2023)	4a	ICCV	2023	96,9
HD-GCN (Lee <i>et al.</i> 2023)	6a	ICCV	2023	97,2
BlockGCN (Zhou <i>et al.</i> 2024)	4a	CVPR	2024	96,9
SelfGCN (Wu <i>et al.</i> 2024)	4a	TIP	2024	96,8
DeGCN (Myung <i>et al.</i> 2024)	4a	TIP	2024	97,2
FCSA-GCN	4a	-	2025	97,2

AGMS-GCN ve FCSA-GCN Mimarilerinin Karmaşıklık Analizi

Tablo 18’de, önerilen yöntemlerin NTU-60 veri kümesinde son teknoloji yöntemlerle doğruluk ve performans açısından karşılaştırmalı analizi sunulmaktadır. Tablo incelendiğinde, doğruluk değerlerinin yanı sıra her yönteme ait parametre sayıları ve FLOPs (Floating Point Operations Per Second) değerleri de verilmiştir. Daha önceki tablolarda yer alan bazı yöntemler

ise Tablo 18’den çıkarılmıştır. Bunun sebebi, bu yöntemlerin orijinal iskelet dizisini tüm 300 kare üzerinden gerçekleştirmek yerine daha kısa bir diziye (örneğin 64 kare) yeniden boyutlandıran veri dönüştürme modülleri içermesidir. Bu durum adil bir karşılaştırmayı engellediğinden, bu yöntemler değerlendirmeye dahil edilmemiştir. Tabloda yer alan * işareti, yöntemlere ait sonuçların yazarlar tarafından yayımlanan kodları çalıştırılarak elde edildiğini belirtmektedir. - işareti ise yönteme ait ilgili değerin yayımlanan makalelerde sağlanmadığını ve kodların çalıştırılması sonucunda da elde edilemediğini göstermektedir.

FLOPs yani saniyedeki kayan nokta işlem sayısı bir modelin hesaplama karmaşıklığını ifade eder. FLOPs değeri, bir modelin belirli bir işlemi gerçekleştirirken kaç tane kayan nokta işlemi yaptığını gösterir. Yüksek FLOPs değeri, modelin daha fazla hesaplama yapması gerektiği anlamına gelir ve genellikle modelin daha fazla bilgi işlediğini veya daha karmaşık yapıya sahip olduğunu gösterir. Örneğin, FC-SA-GCN modelinin 24,4 FLOPs değeri, modelin yaklaşık 24,4 milyar kayan nokta işlemi gerçekleştirdiğini göstermektedir. FLOPs değeri ne kadar yüksekse, modelin hesaplama karmaşıklığı ve hesaplama süresi o kadar uzun olur. FLOPs değeri düşük olan modeller ise daha hızlı çalışır ve daha düşük donanımsal gereksinimleri ihtiyaç duyar. Parametre sayısı ise bir modelin öğrenmesi gereken ağırlık ve önyargı (bias) değerlerinin toplamını ifade eder. Parametreler, modelin girdilerden çıktı üretirken hangi özelliklere daha fazla veya daha az ağırlık vereceğini belirleyen temel bileşenlerdir. Yüksek parametre sayısına sahip modeller, daha fazla bilgi saklayabilir ve daha karmaşık ilişkileri öğrenebilir; ancak bu durum, aynı zamanda daha yüksek bellek gereksinimi ve daha fazla hesaplama maliyeti anlamına gelmektedir.

Tablo 18. Önerilen Yöntemlerin NTU-60 Veri Kümesinde Son Teknoloji Yöntemlerle Doğruluk ve Karmaşıklık Karşılaştırması

Yöntemler	Akış	Yayın	Tarih	NTU-60		Parametre Say. (10 ⁶)	FLOPs (10 ⁹)
				CS	CV		
ST-GCN (Yan <i>et al.</i> 2018)	1a	AAAI	2018	81.5	88.3	3,1*	16,3*
2s-AGCN (Shi <i>et al.</i> 2019b)	2a	CVPR	2019	88.5	95.1	6,9	37,3*
AGC-LSTM (Si <i>et al.</i> 2019)	2a	CVPR	2019	89.2	95.0	22,9	54,4*
DGNN (Shi <i>et al.</i> 2019a)	2a	CVPR	2019	89.9	96.1	26,2	70.0*
AS-GCN (Li <i>et al.</i> 2019)	1a	CVPR	2019	86,8	94.2	9,5*	32,8*
MS-AAGCN (Shi <i>et al.</i> 2020)	4a	TIP	2020	90.0	96.2	15,1	37,4
DynamicGCN (Ye <i>et al.</i> 2020)	4a	ACM	2020	91.5	96.0	14,4	-
MS-G3D (Liu <i>et al.</i> 2020)	2a	CVPR	2020	91.5	96.2	6,4	48,9*
MST-GCN (Chen <i>et al.</i> 2021)	4a	AAAI	2021	91.5	96.6	12,0	-
InfoGCN (Chi <i>et al.</i> 2022)	4a	CVPR	2022	92.7	96.9	6,3	-

FR Head (Zhou <i>et al.</i> 2023)	4a	CVPR	2023	92.8	96.8	8,0	-
HD-GCN (Lee <i>et al.</i> 2023)	4a	ICCV	2023	93.0	97.0	6,7	-
HA-GNN (Geng <i>et al.</i> 2024)	3a	TMM	2024	93.4	97.2	7,1	40,8
AGMS-GCN	4a	KBS	2025	93.9	97.8	15,9	87,9
FCSA-GCN	4a	-	2025	<u>93.6</u>	<u>97.5</u>	5,9	24,4

Tablo 18 incelendiğinde, AGMS-GCN modeli 15,9 milyon parametre ve 87,9 milyar işlem sayısına sahiptir. Bu sonuç, modelin dikkat bitişiklik matrislerinin oluşturulması ve çok ölçekli yapıların ayrı ayrı kullanılması nedeniyle karmaşık bir yapıya sahip olduğunu göstermektedir. Fakat bu karmaşıklık artışı, doğrulukta sağlanan iyileşme ile dengelenmektedir. NTU-60 veri kümesinde CS değerlendirmesinde %93,9 ve CV değerlendirmesinde %97,8 doğruluk elde edilmesi, modelin uzaysal-zamansal ilişkileri etkin bir şekilde öğrenebildiğini ve yüksek genelleme kapasitesine sahip olduğunu ortaya koymaktadır.

FCSA-GCN modeli ise 5,9 milyon parametre ve 24,4 milyar işlem sayısı ile çalışmakta olup, CS değerlendirmesinde %93,6 ve CV değerlendirmesinde %97,5 doğruluk sağlamaktadır. FLOPs ve parametre sayısındaki düşüklüğe rağmen yüksek doğruluk sağlanması, modelin zamansal öz-dikkat mekanizması ve inceden-kabaya özellik çıkarma stratejisinin etkinliğini göstermektedir. Özellikle düşük işlem sayısına rağmen yüksek doğruluk değerlerine ulaşılması, FCSA-GCN modelinin hem hesaplama verimliliğini hem de doğruluk performansını başarılı bir şekilde optimize ettiğini ortaya koymaktadır.

Sonuç olarak, AGMS-GCN modeli, yüksek karmaşıklık ile daha yüksek doğruluk sağlarken, FCSA-GCN modeli ise daha düşük karmaşıklık ve yüksek doğruluk dengesini sağlayarak gerçek dünya uygulamaları için önemli bir avantaj sunmaktadır. Bu durum, AGMS-GCN modelinin doğruluk açısından, FCSA-GCN modelinin ise karmaşıklık açısından mevcut yöntemler arasında en başarılı yöntemler olduğunu açıkça göstermektedir.

SONUÇLAR ve ÖNERİLER

Eylem tanıma, insan hareketlerini ve davranışlarını anlamaya yönelik bir alan olarak hem teknolojik hem de toplumsal birçok alanda önemli bir role sahiptir. İnsan hareketleri, günlük yaşamın ayrılmaz bir parçasıdır ve bu hareketlerin doğru bir şekilde tanımlanması, insan-bilgisayar etkileşiminden sağlık hizmetlerine kadar geniş bir yelpazede uygulama alanı sunmaktadır. Eylem tanıma alanındaki ilk çalışmalarda veri türü olarak RGB videolar kullanılmıştır. Fakat son yıllardaki çalışmalar, aydınlatma değişiklikleri, arka plan karışıklığı ve bakış açısı değişikliklerine karşı daha dayanıklı olması, daha az hesaplama maliyeti gerektirmesi, gizlilik endişelerini azaltması gibi sebeplerden ötürü iskelet verileri üzerinde odaklanmıştır. İskelet verileri, insan vücudunun eklem noktalarını temsil eden verilerdir ve bu veriler, hareketlerin daha soyut ve anlamlı bir şekilde analiz edilmesine olanak tanımaktadır. Bu nedenle, iskelet verileri eylem tanıma alanında giderek daha popüler hale gelmiştir.

Derin öğrenmenin mimarilerinin gelişmesi ile birlikte CNN, RNN, GCN ve Transformer tabanlı yöntemler, iskelet tabanlı eylem tanıma için kullanılan son teknoloji yöntemler haline gelmiştir. Fakat insan iskelet topolojisinin bir çizge olarak ifade edilebilmesi nedeniyle, özellikle uzaysal-zamansal çizge evrimsel ağ yapısı olan ST-GCN'ler, iskelet tabanlı eylem tanıma alanında temel yöntemler haline gelmiştir. ST-GCN, insan iskeletinin topolojik yapısını çizge düğümleri ve kenarları üzerinden ifade ederken, çizge evrim katmanları sayesinde eklemler arasındaki uzaysal-zamansal bağımlılıkları keşfetmektedir. İskelet tabanlı eylem tanıma alanında önemli bir başlangıç noktası olarak kabul edilen ST-GCN, sabit topoloji problemi, uzak mesafeli bağımlılıkların yetersiz temsili, uzun menzilli bağımlılıkların yetersiz temsili ve koordineli hareketlerin yetersiz temsili gibi yapısal sınırlamalara sahiptir.

Bu doktora çalışmasında, ST-GCN modellerinin yapısal sınırlamalarını aşmak için AGMS-GCN ve FCSA-GCN olmak üzere iki farklı ağ yapısı geliştirilmiştir. Önerilen iki mimari yapıda da ilk olarak, ham iskelet verileri etkili bir ön işlem adımından geçmektedir. Ön işlem adımı, iskelet verilerini ötelemeye, kamera bakış açısına ve ölçeğe göre değişmez hale getirerek, farklı bireylerden veya senaryolardan kaynaklanan farklılıkları ortadan kaldırmakta ve verileri standart bir yapıya dönüştürmektedir.

AGMS-GCN mimarisi, dikkat mekanizması rehberliğinde eyleme özgü eklem ilişkilerini temsil eden bitişiklik matrisleri üreterek eklemler arasındaki uzaysal-zamansal

bağımlılıkları daha güçlü şekilde modellemektedir. NTU-60, NTU-120 ve N-UCLA veri kümeleri üzerinde yapılan kapsamlı deneyler, modelimizin rekabetçi ve üstün bir performans ortaya koyduğunu göstermektedir. AGMS-GCN modelinin performansı birkaç önemli faktöre dayanmaktadır: (1) Dikkat mekanizmasının kullanımı, eklemler arasındaki bağımlılıkların gücünü dinamik olarak belirlememizi sağlayarak eylem tanıma performansını iyileştirmiştir. (2) AGMS-GCN modeli çok ölçekli çizge yapıları ile hareket desenlerini farklı ayrıntı düzeylerinde yakalamaktadır. Bu hiyerarşik gösterim, modelin farklı veri kümeleri ve eylem türleri arasında genelleştirilebilir performans sergilemesini sağlamıştır. Genel olarak AGMS-GCN modeli, dikkat mekanizması ve çok ölçekli çizge yapılarının entegrasyonu yoluyla alanı ilerleterek, iskelet tabanlı eylem tanıma görevleri için sağlam ve genelleştirilebilir bir çözüm sunmuştur. Fakat, çok ölçekli çizge yapılarının ayrı ayrı işlenmesi, gerçek zamanlı uygulamalar veya kaynak kısıtlı ortamlarda dağıtım için zorluk oluşturabilecek hesaplama karmaşıklığı getirmektedir. Bu hesaplama karmaşıklığını azaltmak için FCSA-GCN mimarisi geliştirilmiştir.

FCSA-GCN mimarisinde, çok ölçekli özellik çıkarma süreci inceden-kabaya ölçekleme stratejisi kullanılarak tek bir akışta birleştirilmiştir. Bu birleşik çerçevede, ilk olarak en fazla sayıda eklem noktası içeren G_{25} çizge yapısı kullanılarak ince taneli özellik çıkarımı gerçekleştirilmektedir. Bu ayrıntılı yapı sayesinde, ilk aşamada önemli ve detaylı özellik haritaları elde edilmektedir. Ardından, bu haritalar G_{11} ve G_6 çizge yapılarına uygun şekilde havuzlanarak yeni ölçeklerdeki özellik gösterimlerine dönüştürülmektedir. Bu havuzlama işlemleri, koordineli hareketleri daha iyi temsil etmeyi ve sabit topolojinin getirdiği uzaysal sınırlamaları aşmayı amaçlamaktadır. Son olarak, her çizge yapısında ayrı ayrı özellik çıkarımı yapılarak elde edilen özellik haritaları birleştirilmektedir. Her özellik çıkarma işlemi sonrasında, özellik haritalarına zamansal öz-dikkat mekanizması uygulanarak uzun menzilli zamansal bağımlılıkların vurgulanması sağlanmaktadır. Çok ölçekli bu özellik çıkarma süreci, DWA yaklaşımı kullanılarak dengeli bir şekilde optimize edilmektedir. İnceden-kabaya özellik çıkarma stratejisi, çok ölçekli yapıları tek bir akışta birleştirerek mimarinin vücut eklemleri arasındaki hem yerel hem de küresel ilişkileri yakalamasına olanak tanımaktadır. Zamansal öz-dikkat mekanizması ise zamansal bağımlılıkları vurgulayarak daha kapsamlı ve genel bir gösterim üretilmesine yardımcı olmaktadır. Modelin etkinliği NTU-60, NTU-120 ve N-UCLA veri kümeleri üzerinde gerçekleştirilen kapsamlı deneylerle ortaya konmuştur. Geliştirilen mimarilerin her biri, tezin amacı ve katkıları bölümünde verilen araştırma sorularına yanıt sunacak şekilde tasarlanmıştır.

- 1) Uzak eklemeler arasındaki uzaysal bağımlılıkların etkili bir biçimde temsil edilmesi nasıl sağlanabilir?

AGMS-GCN modeli, eklemeler arasındaki uzaysal ilişkileri dinamik bir şekilde temsil etmek amacıyla, dikkat mekanizması rehberliğinde oluşturulan öğrenilebilir bitişiklik matrisleri kullanmaktadır. Bu mekanizma sayesinde, sadece fiziksel olarak komşu eklemeler değil, aynı zamanda uzakta bulunan ancak eylem açısından ilişkili eklemeler arasındaki bağlar da modellenmektedir. Bu yaklaşım, sabit topolojinin ötesine geçerek esnek ve bağlama duyarlı bir uzaysal modelleme sunmaktadır. Ayrıca FCSA-GCN'de uygulanan inceden-kabaya çizge stratejisi, önce en ayrıntılı düzeyde ilişki yakalayıp ardından bu bilgiyi daha az karmaşık yapılarla özetleyerek hem yerel hem küresel bağımlılıkların çok katmanlı temsilini mümkün kılmaktadır.

- 2) Modelin daha bilgilendirici zamansal özelliklere odaklanabilmesi için uzun menzilli zamansal bağımlılıklar nasıl yakalanabilir?

FCSA-GCN mimarisinde, her ölçekte elde edilen özellik haritalarına uygulanan zamansal öz-dikkat mekanizması ile uzun menzilli zamansal bağımlılıkların vurgulanması sağlanmıştır. Bu sayede model, yalnızca zamansal evrişim çekirdek boyutu ile sınırlı bir alanda değil, eylemin tamamı boyunca önemli olan zamansal bağlamları da dikkate alarak daha bilgilendirici temsil üretmiştir. Bu da hareketin zaman içerisindeki evrimini daha doğru şekilde yansıtmakta ve tanıma başarımını artırmaktadır.

- 3) Farklı eklem grupları arasındaki eş zamanlı ve koordineli hareketler nasıl daha etkili bir şekilde modellenabilir?

FCSA-GCN modeli, iskelet yapısını inceden-kabaya ölçekleme stratejisiyle ele alarak önce G_{25} çizge yapısı ile ayrıntılı özellikler çıkarmış, ardından G_{11} ve G_6 çizge yapıları aracılığıyla bu bilgileri daha genel seviyede işlemiştir. Bu hiyerarşik yaklaşım, farklı eklem gruplarının koordineli hareketlerini daha güçlü şekilde temsil etmiş; vücudun bölgesel etkileşimlerini hesaba katarak hareket örüntülerinin daha doğru yakalanmasını sağlamıştır. Benzer şekilde, AGMS-GCN mimarisi de üç farklı ölçekli çizge yapısını bağımsız olarak işleyerek her ölçekteki hareket örüntülerini modellemektedir. Her bir ölçek, farklı seviyelerde hareket ilişkilerini temsil ettiğinden, bu çok ölçekli yapı koordineli hareketlerin hassas biçimde yakalanmasına katkı sağlamıştır.

Önerilen AGMS-GCN ve FCSA-GCN modellerinin işlem adımı ve parametre sayısındaki farklılıklarına rağmen, her iki modelin de doğruluk açısından önceki çalışmalardan daha iyi sonuçlar sunduğu gözlemlenmektedir. AGMS-GCN modeli, karmaşıklığın artmasına

rağmen doğrulukta sağladığı artışla öne çıkarken, FCSA-GCN modeli düşük hesaplama maliyeti ile yüksek doğruluk sağlayarak etkinlik ve verimlilik dengesi sunmaktadır. Bu sonuçlar, önerilen modellerin hem yüksek doğruluk hem de hesaplama verimliliği açısından iskelet tabanlı eylem tanıma mevcut yöntemlerden daha başarılı olduğunu göstermektedir. AGMS-GCN ve FCSA-GCN mimarileri özellikle iskelet tabanlı eylem tanıma için tasarlanmış olsa da temelde çizge yapılı veriler üzerinde çalıştığı için moleküler yapı analizi, sosyal ağ analizi ve trafik tahmini gibi birçok farklı alanda kullanılabilir. Ancak, her bir alana özgü veri yapısı ve problemler dikkate alınarak modellerin uyarlanması gerekmektedir.

AGMS-GCN ve FCSA-GCN modelleri, iskelet tabanlı eylem tanıma görevlerinde başarılar elde etmiştir. Ancak, bu alanda hala geliştirilebilecek birçok yön bulunmaktadır. İleriye dönük çalışmalar için aşağıdaki öneriler sunulabilir:

- Eylem tanıma sürecinde, bir eylemi temsil eden en kritik kareleri belirlenmesi ve bu kareler üzerinde odaklanması hem hesaplama maliyetini azaltabilir hem de modelin performansını artırabilir.
- İskelet verileri, RGB ve derinlik verileri gibi farklı veri türleriyle birleştirilerek daha kapsamlı özellikler çıkarılabilir. Çapraz-dikkat mekanizması kullanılarak farklı modaliteler arasında bilgi alışverişi sağlanabilir. Veri türlerinin entegrasyonu, özellikle karmaşık eylemlerin tanınmasında daha yüksek doğruluk sağlayabilir. Modelin farklı senaryolara adaptasyon kabiliyeti geliştirilebilir.
- İnsan-çevre etkileşimini modellemek için çevresel veriler (örneğin, nesne tespiti veya sahne anlama) ile iskelet verileri birleştirilebilir. Bu yaklaşım, eylem tanıma sistemlerinin daha bağlamsal ve anlamlı sonuçlar üretmesini sağlayabilir.
- Çizge transformer ağ yapıları, çizge yapısına sahip veriler üzerinde çalışan, klasik transformer mimarisinin prensiplerini çizgelere uyarlayan bir derin öğrenme modelidir. Bu ağ yapıları, insan vücudunun topolojik yapısının eylemlere göre değişimlerini modellemek için kullanılabilirse iskelet tabanlı eylem tanıma görevlerinde, eklemler arasındaki uzaysal-zamansal bağımlılıkların daha hassas bir şekilde yakalanmasını mümkün kılabilir.

Bu doğrultuda yapılacak geliştirmeler, iskelet tabanlı eylem tanıma alanındaki ilerlemelere önemli katkılar sağlayacağı düşünülmektedir.

KAYNAKÇA

- Agahian, S., Negin, F., & Köse, C. (2019). Improving bag-of-poses with semi-temporal pose descriptors for skeleton-based action recognition. *The Visual Computer*, 35, 591-607.
- Agahian, S., Negin, F., & Köse, C. (2020). An efficient human action recognition framework with pose-based spatiotemporal features. *Engineering Science and Technology, an International Journal*, 23(1), 196-203.
- Bello, I., Zoph, B., Vaswani, A., Shlens, J., & Le, Q. V. (2019). Attention augmented convolutional networks. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 3286-3295).
- Boulahia, S. Y., Amamra, A., Madi, M. R., & Daikh, S. (2021). Early, intermediate and late fusion strategies for robust deep learning-based multimodal action recognition. *Machine Vision and Applications*, 32(6), 121.
- Calabrese, E., Taverni, G., Awai Easthope, C., Skriabine, S., Corradi, F., Longinotti, L., ... & Delbruck, T. (2019). DHP19: Dynamic vision sensor 3D human pose dataset. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops* (pp. 0-0).
- Cao, C., Lan, C., Zhang, Y., Zeng, W., Lu, H., & Zhang, Y. (2018). Skeleton-based action recognition with gated convolutional neural networks. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(11), 3247-3257.
- Cao, Z., Hidalgo, G., Simon, T., Wei, S. E., & Sheikh, Y. (2019). Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE transactions on pattern analysis and machine intelligence*, 43(1), 172-186.
- Cao, Z., Simon, T., Wei, S. E., & Sheikh, Y. (2017). Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7291-7299).
- Carreira, J., & Zisserman, A. (2017). Quo vadis, action recognition? a new model and the kinetics dataset. In *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 6299-6308).
- Chen, Y., Zhang, Z., Yuan, C., Li, B., Deng, Y., & Hu, W. (2021). Channel-wise topology refinement graph convolution for skeleton-based action recognition. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 13359-13368).
- Chen, Z., Badrinarayanan, V., Lee, C. Y., & Rabinovich, A. (2018, July). Gradnorm: Gradient normalization for adaptive loss balancing in deep multitask networks. In *International conference on machine learning* (pp. 794-803). PMLR.
- Chen, Z., Li, S., Yang, B., Li, Q., & Liu, H. (2021, May). Multi-scale spatial temporal graph convolutional network for skeleton-based action recognition. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 35, No. 2, pp. 1113-1122).
- Cheng, H., & Chung, S. M. (2016). Orthogonal moment-based descriptors for pose shape query on 3D point cloud patches. *Pattern Recognition*, 52, 397-409.
- Cheng, K., Zhang, Y., Cao, C., Shi, L., Cheng, J., & Lu, H. (2020b). Decoupling gcn with dropgraph module for skeleton-based action recognition. In *Computer Vision–ECCV*

2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIV 16 (pp. 536-553). Springer International Publishing.

- Cheng, K., Zhang, Y., He, X., Chen, W., Cheng, J., & Lu, H. (2020a). Skeleton-based action recognition with shift graph convolutional network. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 183-192).
- Chi, H. G., Ha, M. H., Chi, S., Lee, S. W., Huang, Q., & Ramani, K. (2022). Infogcn: Representation learning for human skeleton-based action recognition. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 20186-20196).
- Cho, S., Maqbool, M., Liu, F., & Foroosh, H. (2020). Self-attention network for skeleton-based human action recognition. In Proceedings of the IEEE/CVF winter conference on applications of computer vision (pp. 635-644).
- Do, J., & Kim, M. (2024, September). Skateformer: skeletal-temporal transformer for human action recognition. In European Conference on Computer Vision (pp. 401-420). Cham: Springer Nature Switzerland.
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.
- Du, Y., Fu, Y., & Wang, L. (2015, November). Skeleton based action recognition with convolutional neural network. In 2015 3rd IAPR Asian conference on pattern recognition (ACPR) (pp. 579-583). IEEE.
- Du, Y., Wang, W., & Wang, L. (2015). Hierarchical recurrent neural network for skeleton based action recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1110-1118).
- Duan, H., Zhao, Y., Chen, K., Lin, D., & Dai, B. (2022). Revisiting skeleton-based action recognition. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 2969-2978).
- Elharrouss, O., Almaadeed, N., Al-Maadeed, S., Bouridane, A., & Beghdadi, A. (2021). A combined multiple action recognition and summarization for surveillance video sequences. *Applied Intelligence*, 51, 690-712.
- Feng, D., Wu, Z., Zhang, J., & Ren, T. (2021). Multi-scale spatial temporal graph neural network for skeleton-based action recognition. *IEEE Access*, 9, 58256-58265.
- Feng, L., Zhao, Y., Zhao, W., & Tang, J. (2022). A comparative review of graph convolutional networks for human skeleton-based action recognition. *Artificial Intelligence Review*, 1-31.
- Gao, Y., Xiang, X., Xiong, N., Huang, B., Lee, H. J., Alrifai, R., ... & Fang, Z. (2018). Human action monitoring for healthcare based on deep learning. *Ieee Access*, 6, 52277-52285.
- Geng, P., Lu, X., Li, W., & Lyu, L. (2024). Hierarchical aggregated graph neural network for skeleton-based action recognition. *IEEE Transactions on Multimedia*.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
- Hoshino, S., & Niimura, K. (2019). Optical flow for real-time human detection and action recognition based on CNN classifiers. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 23(4), 735-742.

- Host, K., & Ivašić-Kos, M. (2022). An overview of Human Action Recognition in sports based on Computer Vision. *Heliyon*, 8(6).
- Jiao, Z., Jia, G., & Cai, Y. (2020). Ensuring computers understand manual operations in production: Deep-learning-based action recognition in industrial workflows. *Applied Sciences*, 10(3), 966.
- Ke, L., Peng, K. C., & Lyu, S. (2022, June). Towards to-at spatio-temporal focus for skeleton-based action recognition. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 36, No. 1, pp. 1131-1139).
- Ke, Q., Bennamoun, M., An, S., Sohel, F., & Boussaid, F. (2017). A new representation of skeleton sequences for 3d action recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3288-3297).
- Kendall, A., Gal, Y., & Cipolla, R. (2018). Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7482-7491).
- Kılıç, U., Karadağ, Ö. Ö., & Özyer, G. T. (2024, May). SkelResNet: Transfer learning approach for skeleton-based action recognition. In *2024 32nd Signal Processing and Communications Applications Conference (SIU)* (pp. 1-4). IEEE.
- Kilic, U., Karadag, O. O., & Ozyer, G. T. (2025). AGMS-GCN: Attention-guided multi-scale graph convolutional networks for skeleton-based action recognition. *Knowledge-Based Systems*, 113045.
- Kipf, T. N., & Welling, M. (2016). Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.
- Korban, M., & Li, X. (2020). Ddgcnn: A dynamic directed graph convolutional network for action recognition. In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XX 16* (pp. 761-776). Springer International Publishing.
- Kuehne, H., Jhuang, H., Garrote, E., Poggio, T., & Serre, T. (2011, November). HMDB: a large video database for human motion recognition. In *2011 International conference on computer vision* (pp. 2556-2563). IEEE.
- Lee, I., Kim, D., Kang, S., & Lee, S. (2017). Ensemble deep learning for skeleton-based action recognition using temporal sliding lstm networks. In *Proceedings of the IEEE international conference on computer vision* (pp. 1012-1020).
- Lee, J., Lee, M., Lee, D., & Lee, S. (2023). Hierarchically decomposed graph convolutional networks for skeleton-based action recognition. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 10444-10453).
- Li, C., Hou, Y., Wang, P., & Li, W. (2017). Joint distance maps based action recognition with convolutional neural networks. *IEEE Signal Processing Letters*, 24(5), 624-628.
- Li, C., Sun, S., Min, X., Lin, W., Nie, B., & Zhang, X. (2017, July). End-to-end learning of deep convolutional neural network for 3D human action recognition. In *2017 IEEE international conference on multimedia & expo workshops (ICMEW)* (pp. 609-612). IEEE.
- Li, M., Chen, S., Chen, X., Zhang, Y., Wang, Y., & Tian, Q. (2019). Actional-structural graph convolutional networks for skeleton-based action recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 3595-3603).

- Li, M., Chen, S., Chen, X., Zhang, Y., Wang, Y., & Tian, Q. (2021). Symbiotic graph neural networks for 3d skeleton-based human action recognition and motion prediction. *IEEE transactions on pattern analysis and machine intelligence*, 44(6), 3316-3333.
- Li, X., Chen, H., He, S., Chen, X., Dong, S., Yan, P., & Fang, B. (2023). Action recognition based on multimode fusion for VR online platform. *Virtual Reality*, 27(3), 1797-1812.
- Li, X., He, Y., & Jing, X. (2019). A survey of deep learning-based human activity recognition in radar. *Remote sensing*, 11(9), 1068.
- Li, Y., Xia, R., Liu, X., & Huang, Q. (2019, July). Learning shape-motion representations from geometric algebra spatio-temporal model for skeleton-based action recognition. In 2019 IEEE international conference on multimedia and Expo (ICME) (pp. 1066-1071). IEEE.
- Li, Z., Le Kernec, J., Abbasi, Q., Fioranelli, F., Yang, S., & Romain, O. (2023). Radar-based human activity recognition with adaptive thresholding towards resource constrained platforms. *Scientific Reports*, 13(1), 3473.
- Liang, J., Zhu, H., Zhang, E., & Zhang, J. (2022). Stargazer: A transformer-based driver action detection system for intelligent transportation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 3160-3167).
- Liu, C., Hu, Y., Li, Y., Song, S., & Liu, J. (2017). Pku-mmmd: A large scale benchmark for continuous multi-modal human action understanding. *arXiv preprint arXiv:1703.07475*.
- Liu, J., Shahroudy, A., Perez, M., Wang, G., Duan, L. Y., & Kot, A. C. (2019). Ntu rgb+ d 120: A large-scale benchmark for 3d human activity understanding. *IEEE transactions on pattern analysis and machine intelligence*, 42(10), 2684-2701.
- Liu, J., Shahroudy, A., Xu, D., & Wang, G. (2016, September). Spatio-temporal lstm with trust gates for 3d human action recognition. In *European conference on computer vision* (pp. 816-833). Cham: Springer International Publishing.
- Liu, J., Wang, G., Hu, P., Duan, L. Y., & Kot, A. C. (2017). Global context-aware attention lstm networks for 3d action recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1647-1656).
- Liu, S., Johns, E., & Davison, A. J. (2019). End-to-end multi-task learning with attention. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 1871-1880).
- Liu, Y., Zhang, H., Xu, D., & He, K. (2022). Graph transformer network with temporal kernel attention for skeleton-based action recognition. *Knowledge-Based Systems*, 240, 108146.
- Liu, Z., Zhang, H., Chen, Z., Wang, Z., & Ouyang, W. (2020). Disentangling and unifying graph convolutions for skeleton-based action recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 143-152).
- Loshchilov, I., & Hutter, F. (2017). Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.
- Myung, W., Su, N., Xue, J. H., & Wang, G. (2024). Degcn: Deformable graph convolutional networks for skeleton-based action recognition. *IEEE Transactions on Image Processing*, 33, 2477-2490.
- Ofli, F., Chaudhry, R., Kurillo, G., Vidal, R., & Bajcsy, R. (2013, January). Berkeley mhad: A comprehensive multimodal human action database. In 2013 IEEE workshop on applications of computer vision (WACV) (pp. 53-60). IEEE.

- Pandey, A., & Kumar, P. (2024). Resstanet: deep residual spatio-temporal attention network for violent action recognition. *International Journal of Information Technology*, 16(5), 2891-2900.
- Peng, W., Hong, X., & Zhao, G. (2021). Tripool: Graph triplet pooling for 3D skeleton-based action recognition. *Pattern Recognition*, 115, 107921.
- Piergiovanni, A. J., & Ryoo, M. S. (2019). Representation flow for action recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 9945-9953).
- Plizzari, C., Cannici, M., & Matteucci, M. (2021). Skeleton-based action recognition via spatial and temporal transformer networks. *Computer Vision and Image Understanding*, 208, 103219.
- Prati, A., Shan, C., & Wang, K. I. K. (2019). Sensors, vision and networks: From video surveillance to activity recognition and health monitoring. *Journal of Ambient Intelligence and Smart Environments*, 11(1), 5-22.
- Rani, S. S., Naidu, G. A., & Shree, V. U. (2021). Kinematic joint descriptor and depth motion descriptor with convolutional neural networks for human action recognition. *Materials Today: Proceedings*, 37, 3164-3173.
- Rastgoo, R., Kiani, K., & Escalera, S. (2021). Sign language recognition: A deep survey. *Expert Systems with Applications*, 164, 113794.
- Rodomagoulakis, I., Kardaris, N., Pitsikalis, V., Mavroudi, E., Katsamanis, A., Tsiami, A., & Maragos, P. (2016, March). Multimodal human action recognition in assistive human-robot interaction. In *2016 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 2702-2706). IEEE.
- Sardari, S., Sharifzadeh, S., Daneshkhah, A., Nakisa, B., Loke, S. W., Palade, V., & Duncan, M. J. (2023). Artificial intelligence for skeleton-based physical rehabilitation action evaluation: A systematic review. *Computers in Biology and Medicine*, 158, 106835.
- Sevilla-Lara, L., Liao, Y., Güney, F., Jampani, V., Geiger, A., & Black, M. J. (2019). On the integration of optical flow and action recognition. In *Pattern Recognition: 40th German Conference, GCPR 2018, Stuttgart, Germany, October 9-12, 2018, Proceedings 40* (pp. 281-297). Springer International Publishing.
- Shahrourdy, A., Liu, J., Ng, T. T., & Wang, G. (2016). Ntu rgb+ d: A large scale dataset for 3d human activity analysis. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1010-1019).
- Shi, L., Zhang, Y., Cheng, J., & Lu, H. (2019). Skeleton-based action recognition with directed graph neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 7912-7921).
- Shi, L., Zhang, Y., Cheng, J., & Lu, H. (2019a). Skeleton-based action recognition with directed graph neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 7912-7921).
- Shi, L., Zhang, Y., Cheng, J., & Lu, H. (2019b). Two-stream adaptive graph convolutional networks for skeleton-based action recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 12026-12035).
- Shi, L., Zhang, Y., Cheng, J., & Lu, H. (2020). Decoupled spatial-temporal attention network for skeleton-based action-gesture recognition. In *Proceedings of the Asian conference on computer vision*.

- Shi, L., Zhang, Y., Cheng, J., & Lu, H. (2020). Skeleton-based action recognition with multi-stream adaptive graph convolutional networks. *IEEE Transactions on Image Processing*, 29, 9532-9545.
- Shiraki, K., Hirakawa, T., Yamashita, T., & Fujiyoshi, H. (2020). Spatial temporal attention graph convolutional networks with mechanics-stream for skeleton-based action recognition. In *Proceedings of the Asian Conference on Computer Vision*.
- Si, C., Chen, W., Wang, W., Wang, L., & Tan, T. (2019). An attention enhanced graph convolutional lstm network for skeleton-based action recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 1227-1236).
- Simonyan, K., & Zisserman, A. (2014). Two-stream convolutional networks for action recognition in videos. *Advances in neural information processing systems*, 27.
- Singh, R., Kushwaha, A. K. S., & Srivastava, R. (2019). Multi-view recognition system for human activity based on multiple features for video surveillance system. *Multimedia Tools and Applications*, 78, 17165-17196.
- Song, J., Wang, H., Li, J., Zheng, J., Zhao, Z., & Li, Q. (2025). Hand-aware graph convolution network for skeleton-based sign language recognition. *Journal of Information and Intelligence*, 3(1), 36-50.
- Song, Y. F., Zhang, Z., Shan, C., & Wang, L. (2022). Constructing stronger and faster baselines for skeleton-based action recognition. *IEEE transactions on pattern analysis and machine intelligence*, 45(2), 1474-1488.
- Song, Y., Dai, Y., Jin, T., & Song, Y. (2023). Dual-task human activity sensing for pose reconstruction and action recognition using 4-D imaging radar. *IEEE Sensors Journal*, 23(19), 23927-23940.
- Sun, K., Xiao, B., Liu, D., & Wang, J. (2019). Deep high-resolution representation learning for human pose estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 5693-5703).
- Sun, Z., Ke, Q., Rahmani, H., Bennamoun, M., Wang, G., & Liu, J. (2022). Human action recognition from various data modalities: A review. *IEEE transactions on pattern analysis and machine intelligence*, 45(3), 3200-3225.
- Terreran, M., Barcellona, L., & Ghidoni, S. (2023). A general skeleton-based action and gesture recognition framework for human-robot collaboration. *Robotics and Autonomous Systems*, 170, 104523.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- Vemulapalli, R., Arrate, F., & Chellappa, R. (2014). Human action recognition by representing 3d skeletons as points in a lie group. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 588-595).
- Verma, K. K., Singh, B. M., & Dixit, A. (2022). A review of supervised and unsupervised machine learning techniques for suspicious behavior recognition in intelligent surveillance system. *International Journal of Information Technology*, 14(1), 397-410.
- Wan, S., Xu, X., Wang, T., & Gu, Z. (2020). An intelligent video analysis method for abnormal event detection in intelligent transportation systems. *IEEE Transactions on Intelligent Transportation Systems*, 22(7), 4487-4495.

- Wang, H., & Wang, L. (2017). Modeling temporal dynamics and spatial configurations of actions using two-stream recurrent neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 499-508).
- Wang, J., Nie, X., Xia, Y., Wu, Y., & Zhu, S. C. (2014). Cross-view action modeling, learning and recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2649-2656).
- Wang, L., Ding, Z., Tao, Z., Liu, Y., & Fu, Y. (2019). Generative multi-view human action recognition. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 6212-6221).
- Wang, P., Li, W., Li, C., & Hou, Y. (2018). Action recognition based on joint trajectory maps with convolutional neural networks. *Knowledge-Based Systems*, 158, 43-53.
- Wang, W., Liu, A. X., Shahzad, M., Ling, K., & Lu, S. (2015, September). Understanding and modeling of wifi signal based human activity recognition. In *Proceedings of the 21st annual international conference on mobile computing and networking* (pp. 65-76).
- Wang, W., Liu, A. X., Shahzad, M., Ling, K., & Lu, S. (2017). Device-free human activity recognition using commercial WiFi devices. *IEEE Journal on Selected Areas in Communications*, 35(5), 1118-1131.
- Wang, X., Girshick, R., Gupta, A., & He, K. (2018). Non-local neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7794-7803).
- Wu, F., Wang, Q., Bian, J., Ding, N., Lu, F., Cheng, J., ... & Xiong, H. (2022). A survey on video action recognition in sports: Datasets, methods and applications. *IEEE Transactions on Multimedia*, 25, 7943-7966.
- Wu, Z., Sun, P., Chen, X., Tang, K., Xu, T., Zou, L., ... & Weise, T. (2024). SelfGCN: Graph convolution network with self-attention for skeleton-based action recognition. *IEEE Transactions on Image Processing*.
- Xiao, F., Lee, Y. J., Grauman, K., Malik, J., & Feichtenhofer, C. (2020). Audiovisual slowfast networks for video recognition. *arXiv preprint arXiv:2001.08740*.
- Xie, J., Meng, Y., Zhao, Y., Nguyen, A., Yang, X., & Zheng, Y. (2024, March). Dynamic semantic-based spatial graph convolution network for skeleton-based human action recognition. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 38, No. 6, pp. 6225-6233).
- Xie, J., Miao, Q., Liu, R., Xin, W., Tang, L., Zhong, S., & Gao, X. (2021). Attention adjacency matrix based graph convolutional networks for skeleton-based action recognition. *Neurocomputing*, 440, 230-239.
- Xu, H., Gao, Y., Hui, Z., Li, J., & Gao, X. (2025). Language knowledge-assisted representation learning for skeleton-based action recognition. *IEEE Transactions on Multimedia*.
- Yadav, S. K., Tiwari, K., Pandey, H. M., & Akbar, S. A. (2021). A review of multimodal human activity recognition with special emphasis on classification, applications, challenges and future directions. *Knowledge-Based Systems*, 223, 106970.
- Yan, S., Xiong, Y., & Lin, D. (2018, April). Spatial temporal graph convolutional networks for skeleton-based action recognition. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 32, No. 1).
- Yang, H., Yan, D., Zhang, L., Sun, Y., Li, D., & Maybank, S. J. (2021). Feedback graph convolutional network for skeleton-based action recognition. *IEEE Transactions on Image Processing*, 31, 164-175.

- Yang, L. I., Huang, J., Feng, T. I. A. N., Hong-An, W. A. N. G., & Guo-Zhong, D. A. I. (2019). Gesture interaction in virtual reality. *Virtual Reality & Intelligent Hardware*, 1(1), 84-112.
- Yang, W., Zhang, J., Cai, J., & Xu, Z. (2021). Shallow graph convolutional network for skeleton-based action recognition. *Sensors*, 21(2), 452.
- Yang, X., & Tian, Y. L. (2012, June). Eigenjoints-based action recognition using naive-bayes-nearest-neighbor. In 2012 IEEE computer society conference on computer vision and pattern recognition workshops (pp. 14-19). IEEE.
- Ye, F., Pu, S., Zhong, Q., Li, C., Xie, D., & Tang, H. (2020, October). Dynamic gcn: Context-enriched topology learning for skeleton-based action recognition. In *Proceedings of the 28th ACM international conference on multimedia* (pp. 55-63).
- Zhang, P., Lan, C., Xing, J., Zeng, W., Xue, J., & Zheng, N. (2019). View adaptive neural networks for high performance skeleton-based human action recognition. *IEEE transactions on pattern analysis and machine intelligence*, 41(8), 1963-1978.
- Zhang, P., Lan, C., Zeng, W., Xing, J., Xue, J., & Zheng, N. (2020). Semantics-guided neural networks for efficient skeleton-based human action recognition. In *proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 1112-1121).
- Zhou, H., Liu, Q., & Wang, Y. (2023). Learning discriminative representations for skeleton based action recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 10608-10617).
- Zhou, Y., Yan, X., Cheng, Z. Q., Yan, Y., Dai, Q., & Hua, X. S. (2024). Blockgcn: Redefine topology awareness for skeleton-based action recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 2049-2058).
- Zhu, W., Lan, C., Xing, J., Zeng, W., Li, Y., Shen, L., & Xie, X. (2016, March). Co-occurrence feature learning for skeleton based action recognition using regularized deep LSTM networks. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 30, No. 1).
- Zhu, Y., Shuai, H., Liu, G., & Liu, Q. (2022). Multilevel spatial-temporal excited graph network for skeleton-based action recognition. *IEEE Transactions on Image Processing*, 32, 496-508.
- Ziaeeefard, M., & Ebrahimnezhad, H. (2010, August). Hierarchical human action recognition by normalized-polar histogram. In 2010 20th International Conference on Pattern Recognition (pp. 3720-3723). IEEE.

ÖZGEÇMİŞ

Kişisel Bilgiler	
Adı Soyadı:	Uğur KILIÇ
Doğum tarihi:	
Doğum Yeri:	
Uyruğu:	
Adres:	
Tel:	
E-mail:	
Eğitim	
Lise:	Erzurum Anadolu Lisesi (2004)
Lisans:	Harran Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü (2009)
Yüksek lisans:	Atatürk Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Anabilim Dalı (2015)
Doktora:	Atatürk Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Anabilim Dalı (2025)
Yabancı Dil Bilgisi	
İngilizce:	İyi
Tezden Üretilmiş Yayınlar	
1. Kilic, U., Karadag, O. O., & Ozyer, G. T. (2025). AGMS-GCN: Attention-guided multi-scale graph convolutional networks for skeleton-based action recognition. Knowledge-Based Systems, 113045. 2. Kılıç, U., Karadağ, Ö. Ö., & Özyer, G. T. (2024, May). SkelResNet: Transfer learning approach for skeleton-based action recognition. In 2024 32nd Signal Processing and Communications Applications Conference (SIU) (pp. 1-4). IEEE. 3. Kılıç, U., Karadağ, Ö. Ö., & Özyer, G. T. (2023, October). Experimental Investigation of Spatial-temporal Graph Convolutional Networks and Attentional Mechanism in Skeleton-Based Action Recognition. In 2023 Innovations in Intelligent Systems and Applications Conference (ASYU) (pp. 1-6). IEEE. 4. Fine-to-Coarse Self-Attention Graph Convolutional Network for Skeleton-Based Action	

Tez Dışı Üretilmiş Yayınlar

1. Yıldırım Demirdöğen, E., Akıncı, M. A., Bozkurt, A., Turan, B., Esin, İ. S., Donbaloğlu, M. A., ... & Dursun, O. B. (2024). Objective Parameters in Attention Deficit Hyperactivity Disorder: Eye and Head Movements. *Journal of Attention Disorders*, 28(6), 982-991.
2. Kilic, U., Karabey Aksakalli, I., Tumuklu Ozyer, G., Aksakalli, T., Ozyer, B., & Adanur, S. (2023). Exploring the Effect of Image Enhancement Techniques with Deep Neural Networks on Direct Urinary System X-Ray (DUSX) Images for Automated Kidney Stone Detection. *International Journal of Intelligent Systems*, 2023(1), 3801485.
3. Kılıç, U., Demir, Y., Kaplan, N. H., & Tümüklü Özyer, G. (2021). Recognition of Electronic Circuit Components Using a Modern Convolutional Object Detector. In 1st International Conference on Advances in Engineering, Architecture, Science and Technology.