



T.C.
NİĞDE ÖMER HALİSDEMİR ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ELEKTRİK-ELEKTRONİK MÜHENDİSLİĞİ ANA BİLİM DALI

DERİN AĞ TABANLI İYİLEŞTİRİLMİŞ STEREO EŞLEŞTİRME

ALPER EMLEK

Ağustos 2022

T.C.
NİĞDE ÖMER HALİSDEMİR ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ELEKTRİK-ELEKTRONİK MÜHENDİSLİĞİ ANA BİLİM DALI

DERİN AĞ TABANLI İYİLEŞTİRİLMİŞ STEREO EŞLEŞTİRME

ALPER EMLEK

Doktora Tezi

Danışman

Dr. Öğr. Üyesi Murat PEKER

Ağustos 2022

Alper EMLEK tarafından **Dr. Öğr. Üyesi Murat PEKER** danışmanlığında hazırlanan “**DERİN AĞ TABANLI İYİLEŞTİRİLMİŞ STEREO EŞLEŞTİRME**” adlı bu çalışma jürimiz tarafından Niğde Ömer Halisdemir Üniversitesi Fen Bilimleri Enstitüsü **Elektrik-Elektronik Mühendisliği** Ana Bilim Dalı’nda Doktora tezi olarak kabul edilmiştir.

Başkan : Prof. Dr. Erkan ÜLKER, Konya Teknik Üniversitesi

Üye : Doç. Dr. Gökhan GELEN, Bursa Teknik Üniversitesi

Üye : Dr. Öğr. Üyesi Murat PEKER, Bursa Teknik Üniversitesi

Üye : Dr. Öğr. Üyesi Mehmet Kürşat YALÇIN, Niğde Ömer Halisdemir Üniversitesi

Üye : Dr. Öğr. Üyesi Salim ÇINAR, Niğde Ömer Halisdemir Üniversitesi

ONAY:

Bu tez, Fen Bilimleri Enstitüsü Yönetim Kurulunca belirlenmiş olan yukarıdaki jüri üyeleri tarafından/...../20.... tarihinde uygun görülmüş ve Enstitü Yönetim Kurulu’nun/...../20.... tarih ve sayılı kararıyla kabul edilmiştir.

...../...../20...

Prof. Dr. Murat GÖKÇEK
MÜDÜR V.

TEZ BİLDİRİMİ

Tez içindeki bütün bilgilerin bilimsel ve akademik kurallar çerçevesinde elde edilerek sunulduğunu, ayrıca tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.

Alper EMLEK

ÖZET

DERİN AĞ TABANLI İYİLEŞTİRİLMİŞ STEREO EŞLEŞTİRME

EMLEK, Alper

Niğde Ömer Halisdemir Üniversitesi

Fen Bilimleri Enstitüsü

Elektrik-Elektronik Mühendisliği Anabilim Dalı

Danışman : Dr. Öğr. Üyesi Murat PEKER

Ağustos 2022, 103 sayfa

Bu tezde, stereo görüntüler kullanılarak gerçek zamanlı eşitsizlik haritaları elde etmek ve eşitsizlik haritalarını düzeltmek amacıyla çeşitli derin öğrenme tabanlı ağ yapıları önerilmiştir. Tez çalışması kapsamında ilk olarak eşitsizlik haritalarının başarısını artırmaya yönelik ağ yapılarının tasarımları gerçekleştirilmiştir. Bu çalışmalarda eşitsizlik haritalarının düzeltilmesi ve eşleşme maliyetlerinin doğruluğunun artırılması amaçlanmıştır. Sonraki çalışmalarda ise uçtan uca öğrenmeye dayalı tasarımlarla eşitsizlik haritalarının elde edilmesi amaçlanmıştır. Uçtan uca ağ yapıları ile yapılan çalışmaların ilkinde evrişim işlemlerinde kullanılan çekirdek yapılarının stereo eşleştirme problemlerine göre özelleştirilmesiyle bir ağ yapısı oluşturulmuştur. Diğer bir ağ yapısında ise baştan sona öz tasarım ile gerçekleştirilen uçtan uca öğrenme ile eşitsizlik haritası çıkarabilen bir ağ yapısı oluşturulmuştur. Bu ağ yapısında öznitelik çıkarma ve eşleşme maliyetlerinin toplanması adımları için alt ağ yapılarıda önerilmiştir. Ayrıca önerilen bu yöntemlerin tasarımında gerçek zamanlı uygulamalar için uygun çalışma sürelerinin elde edilmesi öncelikli hedef olarak belirlenmiştir. Önerilen yapıların farklı veri kümeleri üzerinde eğitim ve test işlemleri gerçekleştirilmiş olup hem nitel hemde nicel karşılaştırmalar gerçekleştirilerek başarımları değerlendirilmiştir.

Anahtar Sözcükler: Stereo eşleştirme, eşitsizlik haritası, eşitsizlik haritası iyileştirme, derin öğrenme, evrişimsel sinir ağları, yinelenen sinir ağları

SUMMARY

ENHANCED STEREO MATCHING BASED ON DEEP NETWORKS

EMLEK, Alper

Niğde Ömer Halisdemir University

Graduate School of Natural and Applied Sciences

Department of Electrical and Electronics Engineering

Supervisor

: Assistant Professor Dr. Murat PEKER

August 2022, 103 pages

In this thesis, various deep learning-based network structures have been proposed to obtain real-time disparity maps using stereo images and to refine disparity maps. Firstly, the accuracy of disparity maps is increased by designing new network structures. In these designs, it is aimed to refine the disparity maps and increase the accuracy of the matching costs. In the following studies, it is aimed to obtain disparity maps with end-to-end learning-based designs. In the designed first end-to-end network structure, a network structure is created by customizing the kernel structures used in convolution operations according to stereo matching problems. In the second structure, a self-designed network structure has been designed to obtain disparity maps. Here, to extract features and aggregate matching costs, different sub-modules are proposed. In addition, the suitable run time for real-time applications is the top priority while designing these proposed methods. Training and testing processes of the proposed structures are carried out on various datasets, and their performance is evaluated by performing both qualitative and quantitative comparisons.

Keywords: Stereo matching, disparity map, disparity map refinement, deep learning, convolutional neural networks, recurrent neural networks

ÖN SÖZ

Bu doktora tezi çalışmasında, stereo görüntüler kullanılarak eşitsizlik haritaları elde etmek ve eşitsizlik haritalarını düzeltmek amacıyla çeşitli derin öğrenme tabanlı tasarımlar gerçekleştirilmiştir. Önerilen algoritmanın başarımı, farklı veri kümeleri üzerinde test edilmiştir ve diğer algoritmalarla karşılaştırılmıştır.

Bu tez çalışmasının hazırlanmasında bilgi, tecrübe ve desteğini benden esirgemeyen danışmanım Sayın Dr. Öğr. Üyesi Murat PEKER'e en içten teşekkürlerimi sunarım. Doktora çalışmam süresince katkıları ve yön verici yorumlarından dolayı değerli hocalarım Sayın Prof. Dr. Erkan ÜLKER ve Dr. Öğr. Üyesi Mehmet Kürşat YALÇIN'a teşekkürlerimi sunarım.

Bu tezi, bu günlere gelmemde maddi ve manevi her türlü desteğini esirgemeyen babam Rahim EMLEK'e, annem Müjgân EMLEK'e, kardeşim Aylin MUTLU'ya ve özellikle her zaman yanımda olan hayat arkadaşım Betül OSKAYBAŞ EMLEK'e ithaf ediyorum.

Bu çalışma, MMT 2019/7-BAGEP numaralı “Stereo Görüntü Çiftlerinden Derin Öğrenme Tabanlı Gerçek Zamanlı Derinlik Bilgisi Çıkarımı” isimli BAP projesinden üretilmiş olup, projeye destek sağlayan Niğde Ömer Hlaidemir Üniversitesi Bilimsel Araştırma Projeleri Birimine katkılarından dolayı teşekkür ederim ve doktora çalışmam boyunca BİDEB 2211-A Genel Yurt İçi Doktora Burs Programı'yla destekleyen TÜBİTAK'a teşekkür ederim.

İÇİNDEKİLER

ÖZET	iv
SUMMARY	v
ÖN SÖZ	vi
İÇİNDEKİLER	vii
ÇİZELGELER DİZİNİ	ix
ŞEKİLLER DİZİNİ	xi
SİMGE VE KISALTMALAR	xiii
BÖLÜM I GİRİŞ	1
1.1 Stereo Eşleştirme Yöntemleri	2
1.1.1 Geleneksel yöntemler	3
1.1.2 Derin öğrenme tabanlı yöntemler	6
1.1.2.1 CNN tabanlı eşleşme maliyeti hesaplama	8
1.1.2.2 CNN tabanlı aykırı değer tespiti ve fark haritası düzeltme	9
1.1.2.3 CNN tabanlı uçtan-uca öğrenme	12
1.2 Tez Kapsamında Gerçekleştirilen Çalışmalar	16
BÖLÜM II PROBLEM, VERİ VE SİMÜLASYON	18
2.1 Genel Değerlendirmeler	18
2.2 Veri Kümeleri	19
2.3 Performans Ölçütleri	21
2.4 Nicel ve Nitel Karşılaştırmalar	24
2.5 Deney Sistemi	25
BÖLÜM III EŞLEŞME MALİYETSİZ EŞİTSİZLİK HARİTASI DÜZELTME YÖNTEMİ	26
3.1 Ağ Yapısı	26
3.2 Deneysel Sonuçlar	29
BÖLÜM IV YİNELENEN SİNİR AĞI TABANLI EŞLEŞTİRME MALİYETİ İYİLEŞTİRME AĞI	34
4.1 Ağ Yapısı	34
4.2 Deneysel Sonuçlar	38
4.2.1 Stereo metotlar	38
4.2.2 Veri kümeleri ve eğitim detayları	39

4.2.3 Parametre eniyileme	43
4.2.4 Nicel karşılaştırmalar	45
4.2.5 Nitel karşılaştırmalar	48
BÖLÜM V YATAY DİKKATLİ EŞİTSİZLİK AĞI	54
5.1 Ağ Yapısı	55
5.2 Deneysel Sonuçlar	57
BÖLÜM VI PARALEL PİRAMİT HAVUZU TABANLI STEREO AĞI	61
6.1 Ağ Yapısı	62
6.1.1 Başlangıç özniteliklerinin çıkarımı	63
6.1.2 Paralel piramit havuzu yapısı	63
6.1.3 Eşleşme maliyetleri	66
6.1.4 Eşitsizlik haritasının elde edilmesi	67
6.1.5 Kayıp fonksiyonu	68
6.2 Deneysel Sonuçlar	70
6.2.1 Ablasyon çalışmaları	71
6.2.2 Scene Flow veri kümesinde sonuçların karşılaştırma	76
6.2.3 KITTI veri kümelerinde sonuçların karşılaştırması	80
BÖLÜM VII SONUÇ	85
KAYNAKLAR	88
ÖZ GEÇMİŞ	102
TEZ ÇALIŞMASINDAN ÜRETİLEN ESERLER	103

ÇİZELGELER DİZİNİ

Çizelge 2.1. Literatürde en çok kullanılan veri kümeleri ve özellikleri	19
Çizelge 3.1. SimRef ağ yapısında farklı CNN katman sayıları kullanılarak elde edilen 3 piksel hata oranları	30
Çizelge 4.1. KITTI 2012 veri kümesinin farklı bölme oranlarında üç piksel test hata oranları	42
Çizelge 4.2. KITTI 2015 veri kümesinin farklı bölme oranlarında üç piksel test hata oranları	43
Çizelge 4.3. Zaman adımı değeri sabit tutularak, farklı gizli katman değerleri için elde edilen ağ performans değerleri	44
Çizelge 4.4. CR-RNN'nin farklı zaman adımı değerlerinde KITTI 2012 test seti üzerindeki başarımı	44
Çizelge 4.5. CR-RNN'nin farklı zaman adımı değerlerinde KITTI 2015 test seti üzerindeki başarımı	45
Çizelge 4.6. Ağ tabanlı yöntemlerin parametre boyutlarının karşılaştırılması	46
Çizelge 4.7. CR-RNN ağ yapısının KITTI 2012 test setinde farklı stereo metotlarla karşılaştırılması	47
Çizelge 4.8. CR-RNN ağ yapısının KITTI 2015 test setinde farklı stereo metotlarla karşılaştırılması	47
Çizelge 4.9. KITTI 2012 ve KITTI 2015 çevrimiçi test kümelerinde orijinal giriş metotları ve CR-RNN _P ağ yapısı ile elde edilen (%) hata oranlarının karşılaştırılması	48
Çizelge 5.1. HADN, DispNetC ve diğer yöntemlerin Scene Flow veri kümesinde EPE hata oranının ve çalışma sürelerinin karşılaştırılması	58
Çizelge 6.1. Başlangıç özniteliklerinin çıkarılmasında kullanılan alt ağ yapısına ait katman giriş/çıkış topolojisi	63
Çizelge 6.2. Üç katmalı paralel piramit havuzu yapısında kullanılan alt ağ yapısına ait katman giriş/çıkış topolojisi	65
Çizelge 6.3. Paralel piramit havuzundan elde edilen özniteliklere uygulanan korelasyon işlemleri	66
Çizelge 6.4. Eşitsizlik haritalarının elde edilmesinde kullanılan alt ağ yapısına ait	

katman giriş/çıkış topolojisi	69
Çizelge 6.5 P3SNet yapısındaki P3 alt katmanının farklı ölçek değerleri için ablasyon çalışması sonuçları	72
Çizelge 6.6 Çift doğrusal aradeğerleme kullanılarak elde edilen farklı ölçekteki ağların ablasyon çalışması sonuçları	74
Çizelge 6.7 Öznitelik çıkarma katmanında evrimsel sinir ağlarının çekirdek boyutları üzerinde yapılan ablasyon çalışması sonuçları	75
Çizelge 6.8. Farklı ölçeklerdeki eşitsizlik haritalarının ortalama EPE hataları ve elde edilme süreleri	76
Çizelge 6.9. Scene Flow test setinde performans karşılaştırma sonuçları	78
Çizelge 6.10. KITTI 2012 çevrimiçi test setinden elde edilen karşılaştırma sonuçları ..	81
Çizelge 6.11. KITTI 2015 çevrimiçi test setinden elde edilen karşılaştırma sonuçları ..	82

ŞEKİLLER DİZİNİ

Şekil 1.1. Sol ve sağ kameradan elde edilen görüntüler ve bu görüntülerden elde edilmiş eşitsizlik haritası	2
Şekil 2.1. KITTI 2012 veri kümesindeki eğitim kümesinden örnek görüntüler	20
Şekil 2.2. KITTI 2015 veri kümesindeki eğitim kümesinden örnek görüntüler	20
Şekil 2.3. Scene Flow veri kümesinde yer alan eğitim kümesinden örnek görüntüler ...	22
Şekil 3.1. Eşitsizlik haritalarının düzeltilmesi amacıyla önerilen SimRef ağ yapısı	27
Şekil 3.2. Content-CNN ağ yapısı	28
Şekil 3.3. KITTI 2015 veri kümesinde Content-CNN ve SimRef ağ yapısı kullanılarak elde edilen eşitsizlik haritalarının karşılaştırılması	32
Şekil 4.1. Tipik bir RNN yapısı	35
Şekil 4.2. RNN hücresi	36
Şekil 4.3. Yinelenen sinir ağı tabanlı maliyet iyileştirme ağı (Cost Refinement Recurrent Neural Network - CR-RNN) yapısı	36
Şekil 4.4 KITTI 2012 ve KITTI 2015 veri kümelerinde yer alan eğitim setlerindeki eşitsizlik değerlerinin dağılımı	41
Şekil 4.5. KITTI 2015 test seti üzerinde veri artırma işlemi yapılarak ve yapılmadan elde edilen sonuçlarda eşitsizlik dağılımı	42
Şekil 4.6. KITTI 2012 ve KITTI 2015 test veri kümelerindeki ARD eğrileri	49
Şekil 4.7 CR-RNN ve MC-CNN yöntemlerinin KITTI 2012 test veri kümesi kullanarak elde ettikleri eşitsizlik ve hata haritaları	51
Şekil 4.8. CR-RNN ve MC-CNN yöntemlerinin KITTI 2015 test veri kümesi kullanarak elde ettikleri eşitsizlik ve hata haritaları	51
Şekil 4.9. CR-RNN ve CBCA yöntemlerinin KITTI 2012 test veri kümesi kullanarak elde ettikleri eşitsizlik ve hata haritaları	52
Şekil 4.10. CR-RNN ve CBCA yöntemlerinin KITTI 2015 test veri kümesi kullanarak elde ettikleri eşitsizlik ve hata haritaları	52
Şekil 5.1. Yatay dikkatli evrişim (HAC) yapısı	55
Şekil 5.2. Yatay dikkatli eşitsizlik ağı (HADN)	57
Şekil 5.3. Scene Flow veri kümesi üzerindeki nitel karşılaştırmalar	59
Şekil 6.1. Paralel piramit havuzu tabanlı stereo ağ yapısı	62
Şekil 6.2. Paralel piramit havuzu yapısı	64

Şekil 6.3. Farklı ölçek değerlerine sahip ağlar tarafından üretilen eşitsizlik haritaları...	73
Şekil 6.4. Farklı boyutlarda elde edilen eşitsizlik ve hata haritaları	77
Şekil 6.5. Scene Flow veri kümesi üzerindeki nitel karşılaştırmalar	79
Şekil 6.6. KITTI 2012 çevrimiçi test setindeki nitel karşılaştırmalar	83
Şekil 6.7. KITTI 2015 çevrimiçi test setinde ki itel karşılaştırmalar	83



SİMGE VE KISALTMALAR

Simgeler	Açıklama
b	Taban uzunluğu
B	Kötü eşleşen piksel hata oranlarını
d	Eşitsizlik
d_e	Tahmin edilen eşitsizlik değerini
d_g	Doğru eşitsizlik değerini
f	Odak uzaklığı
g	Genişlik
k	Örnekleme sayısı
K	Toplam örnek sayısı
R_k	Ölçüm aralığı
y	Yükseklik
y_e	Tahmin edilen derinlik değerini
y_g	Doğru derinlik değerini
Z	Derinlik
δ	Hata eşiği

Kısaltmalar	Açıklama
1B	Bir Boyutlu
2B	İki Boyutlu
3B	Üç Boyutlu
4B	Dört Boyutlu
AP	Ortalama Havuzlama
ARD	Mutlak Bağlı Fark
ASW	Uyarlanabilir Destek Ağırlığı
BF	Bilateral Filtre
BG	Arka Plan

BI	Çift Doğrusal Aradeğerleme
BP	İnanç Yayılımı
BN	Yığın Normalizasyonu
CBCA	Çapraz Tabanlı Maliyet Toplama
CE	Çapraz Entropi
CNN	Evrişimsel Sinir Ağı
CPU	Merkezi İşlem Birimi
DP	Dinamik Programlama
ESA	Evrişimsel Sinir Ağı
EPE	Uç Nokta Hatası
FC	Tam Bağlı
FG	Ön Plan
FPGA	Alan Programlanabilir Kapı Dizileri
GC	Çizge Kesim
GD	Küresel Fark
GF	Kılavuzlu Filtre
GPU	Grafik İşlemci Birimi
GRU	Kapılı Tekrarlayan Birimler
HAC	Yatay Dikkatli Evrişim
HADN	Yatay Dikkatli Eşitsizlik Ağı
HDA	Hiyerarşik Eşitsizlik Toplama
L1	Manhattan Mesafesi
LGA	Yerel GÜdümlü Toplama
LSTM	Uzun-Kısa Süreli Bellek
LRC	Sol-Sağ Tutarlılık Kontrolü
MF	Medyan Filtresi
MP	En Büyük Havuzlama
MRF	Markof Rastgele Alanı
NAS	Sinir Mimarisi Arama
NCC	Normalleştirilmiş Çapraz Korelasyon
NIG	Normal Ters-Gama
P3	Paralel Piramit Havuzu
P3SNET	Paralel Piramit Havuzu Tabanlı Stereo Ağı
PPP	Paralel Piramit Havuzu

PSMNet	Piramit Stereo Eşleştirme Ağı
PX	Piksel
ReLU	Doğrultulmuş Doğrusal Ünite
RNN	Yinelenen Sinir Ağı
SAD	Mutlak Farklar Toplamı
SIFT	Ölçek Bağımsız Öznitelik Dönüşümü
SLAM	Eşzamanlı Konum Belirleme ve Harita Oluşturma
SPP	Uzaysal Piramit Havuzu
SSD	Karesel Farklar Toplamı
SGD	Rasgele Gradyan İnişi
SGM	Yarı Küresel Eşleşme
ToF	Uçuş Süresi
WCE	Ağırlıklı Çapraz Entropi
WMF	Ağırlıklı Medyan Filtresi
WTA	Kazanan-Hepsini-Alır

BÖLÜM I

GİRİŞ

Son yıllarda, üç boyutlu (3B) verilerin mevcudiyeti ile büyük ölçekli sahnelerin yeniden yapılandırması (Shen, 2013; Michailidis vd.,2014), SLAM (Simultaneous Localization and Mapping, Eşzamanlı Konum Belirleme ve Harita Oluşturma) (Yeh ve Lin, 2021), nesne takibi (Tang ve Hwang2019; Marshall vd., 2021), otonom sürüş (Sun vd., 2019; Weon vd.,2020; Zhao vd., 2020), robotik (Guo vd., 2019) gibi çok sayıda üst düzey bilgisayarlı görü alanında ilerleme kaydedilmiştir. Bu nedenle, doğru ve yoğun derinlik bilgisinin elde edilmesi, bu uygulamaların başarısı için çok önemli bir adımdır.

3B veriler ise sensörler ve merkezi işlem birimi (Central Process Unit, CPU), grafik işlemci birimi (Graphics Processing Unit, GPU), alan programlanabilir kapı dizileri (Field-Programmable Gate Array, FPGA) gibi donanımlar kullanılarak 3B algılama teknikleri ile elde edilmektedir. 3B algılama teknikleri aktif ve pasif teknolojiler olarak sınıflandırılmaktadır. Aktif teknolojilerde derinlik bilgisi lazer, radar gibi bir ışık veya radyo dalgası gibi yardımcı kaynaklar ile elde edilmektedir. Aktif teknolojiler (örn., Kinect, vb...) iç mekan sahnelerinde oldukça iyi performans gösterirken, bazı sınırlamalar nedeniyle genellikle dış ortamlarda düşük performans göstermektedirler. Örneğin, Kinect gibi aktif sensörlerin sınırlı bir çalışma aralığı vardır ve güneş ışığı veya aynı alanı algılayan birden fazla cihazın varlığında etkinlikleri büyük ölçüde düşmektedir.

Pasif teknolojiler ise kamera veya kameralar yardımı ile toplanan görüntülerde piksel eşleştirmeleri yaparak derinlik bilgisi elde etmektedir. Pasif teknolojilerden olan stereo ise aktif teknolojilerin sorun yaşadığı senaryolarda oldukça iyi performans göstermektedir. Ayrıca, stereo gerçek zamanlı derinlik bilgisi elde etmek için uygundur ve daha kapsamlı tekniklere kıyasla düşük maliyet ve gereksinimleri sayesinde çok çeşitli senaryolarda kullanılabilir.

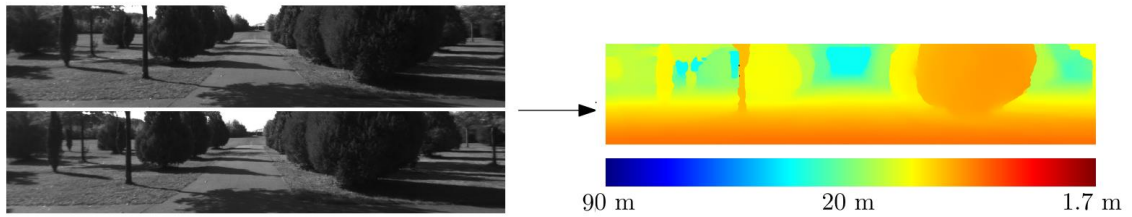
Stereo yöntemlerinde, iki yatay kameradan alınan eş zamanlı görüntüler düzeltildikten sonra giriş olarak kabul edilmektedir. Düzeltme işlemlerinde görüntülerdeki lens bozulmalarının giderilmesinin yanı sıra her her iki görüntüde tespit edilen öznitelikler aynı yatay eksne getirilmektedir. Bu sayede görüntülerde, sol görüntüde yer alan (x, y)

konumundaki pikselin sağ görüntüdeki karşılığını yatay bir arama düzlemi içerisinde sol görüntüde tespit etmek amaçlanmaktadır. Stereo eşleştirme yönteminde, iki görüntüde eşleşen piksellerin, yatay eksenleri arasındaki fark eşitsizlik (disparity) değeri olarak adlandırılır ve d ile gösterilmektedir. Bu durumda, eşleştirilen pikselin sağ görüntüdeki konumu $(x - d, y)$ ile ifade edilmektedir. Eşitsizlik değeri elde edildikten sonra, bu piksele karşılık gelen Z derinlik değeri Denklem (1.1) kullanılarak hesaplanmaktadır.

$$Z = f (b/d) \quad (1.1)$$

Denklem (1.1)'de verilen terimlerden f kameranın odak uzaklığını (focal length) ve b terimi ise iki kamera merkezi arasındaki mesafeyi (baseline length) temsil etmektedir.

Eşitsizlik değeri görüntülerdeki her piksel değeri için hesaplandığı durumda elde edilen harita, eşitsizlik haritası olarak adlandırılmaktadır. Şekil 1.1'de sol ve sağ görüntüden elde edilen örnek bir eşitsizlik haritası verilmektedir.



Şekil 1.1. Sol ve sağ kameradan elde edilen görüntüler ve bu görüntülerden elde edilmiş eşitsizlik haritası (Žbontar ve Lecun, 2015)

Derin öğrenme alanındaki çalışmaların hız kazanması stereo eşleştirme alanında da birçok çözüm yöntemini beraberinde getirmektedir. Bu çözüm yöntemleri ile elde edilen sonuçlar geleneksel metotlardan elde edilen sonuçlardan daha başarılıdır. Dolayısıyla son zamanlarda stereo eşleştirme alanındaki çalışmalar ağırlıklı olarak derin öğrenme teknikleri ile eşitsizlik haritası üretme yöntemleri üzerine yoğunlaşmaktadır.

1.1 Stereo Eşleştirme Yöntemleri

Stereo eşleştirme yöntemlerinde, aynı eksen üzerinde konumlandırılmış iki kameradan alınan eş zamanlı görüntüler kullanılmaktadır. Genellikle stereo kameraların yatay düzlemde yerleştirildiği kurulumlar tercih edilmektedir. Böylece sağ ve sol görüntüler

stereo görüntü çifti olarak kullanılmaktadır. Bu görüntü çiftlerine, ilk olarak, eksen düzeltme işlemi uygulanmaktadır. Düzeltme işlemi, iki farklı yöntem kullanılarak gerçekleştirilebilmektedir. Bunlardan ilkinde satranç tahtası gibi şekli ve ölçüleri bilinen bir nesne yardımıyla kamera parametreleri ve iki kameranın birbirine göre konumlarının belirlenmesi ile yapılmaktadır. (J. Chen vd., 2012). Diğer yöntemde ise her iki kameradan aynı anda alınan görüntülerde belirlenen özniteliklerin aynı doğru üzerine getirilmesi ile elde edilmektedir (Lim vd., 2004).

Bu tez kapsamında düzeltilmiş stereo görüntüden eşitsizlik haritası elde etmek amacıyla kullanılan yöntemler geleneksel ve derin öğrenme tabanlı yöntemler olmak üzere iki gruba ayrılmıştır ve bu başlıklar alt bölümlerde incelenmiştir.

1.1.1 Geleneksel yöntemler

Literatür, stereo eşleştirme problemini ele alan çalışmalar açısından oldukça zengindir. Scharstein ve Szeliski (2002) çalışmalarında, stereo eşleştirme işlemlerindeki adımları daha iyi tanımlayabilmek için dört adımdan oluşan bir yapı taşı önermişlerdir. Bu yapı taşına göre stereo algoritmalar genellikle aşağıdaki dört adımı takip etmektedir:

1. Eşleşme maliyetlerini hesaplama,
2. Maliyetleri toplama,
3. Eşitsizlik hesaplama/eniyileme,
4. Eşitsizlik haritası düzeltme.

Scharstein ve Szeliski'ye (2002) göre stereo eşleştirme algoritmaları eniyileme yaklaşımlarına göre yerel (loca) ve küresel (global) yöntemler olmak üzere ikiye ayrılmaktadır. Yerel yöntemler (Zhang vd. 2009; Hosni vd., 2009; Hosni vd., 2011) tipik olarak bu dört adım içerisinde 1. ve 2. adıma odaklanırken, küresel yöntemler (Mattoccia vd., 2007; Geiger vd., 2010) ise çoğunlukla 1. ve 3. adıma odaklanmaktadır.

Küresel yöntemler, eşitsizlik haritasını belirlemek için tanımladıkları bir enerji fonksiyonunu en aza indirmeyi amaçlamaktadırlar. Küresel yöntemlerde enerji fonksiyonunun genel tanımlaması Denklem (1.2)'de verildiği gibi yapılmaktadır.

$$E(d) = E_{veri}(d) + \lambda E_{düzlük}(d) \quad (1.2)$$

Literatürde verilen çalışmalarda küresel yöntemlerde enerji fonksiyonu tanımlandıktan sonra, tanımlanan enerji fonksiyonunun minimum noktalarını bulmak amacı ile birçok eniyileme algoritması kullanılmıştır. Bu eniyileme algoritmalarından en çok kullanılanları, çizge kesim (Graph Cut, GC) (Kolmogorov ve Zabih, 2006; Bleyer ve Gelautz, 2007; Lu vd., 2021), dinamik programlama (Dynamic Programming, DP) (Ohta ve Kanade, 1985; Forstmann vd., 2004; Kim vd., 2005) ve inanç yayılımıdır (Belief Propagation, BP) (Guan vd., 2009; J. Sun vd., 2003).

Yerel yöntemler ise önceden tanımlanmış parametrelere göre sabit bir şekle sahip belirli bir destek penceresi veya her bir piksel için çevresindeki komşu piksellere göre şekli değişebilen bir destek penceresi kullanmaktadır. Sol görüntü ile sağ görüntü arasında her bir piksel değeri için bu belirlenen destek pencereleri kullanılmakta ve karşılaştırmalar yapılarak eşleşme maliyetleri hesaplanmaktadır.

Piksel değerlerinin karşılaştırılmasında kullanılan temel yöntemler incelendiğinde, iki piksel arasındaki farkın oluşabilecek farklı ortam koşullarında bile en iyi şekilde ifade edilmesi amaçlanmaktadır. Bu yöntemlerden bazıları; mutlak farkların toplamı (Sum of Absolute Differences, SAD) (Perri vd., 2006; Ambrosch vd., 2007; Hamzah vd., 2010), kare farkların toplamı (Sum of Squared Differences, SSD) (Okutomi ve Kanade, 1993; Shi vd., 2001), normalleştirilmiş çapraz korelasyon (Normalized Cross-Correlation, NCC) (Zhang vd., 2009; Heo vd., 2010; Dinh vd., 2016) ve census (Banks vd., 1997; Zinner vd., 2008) dönüşümüdür. Bazı yöntemler ise bu metotlardan bir kaçını belirli ağırlık oranları ile birleştirerek daha iyi temsil özelliği katmayı amaçlamışlardır (Fookes vd., 2002; Hirschmuller, 2005; Hirschmuller ve Scharstein 2007; Ernst ve Hirschmüller, 2008).

Stereo eşleştirme işlemlerinde ikinci aşama olan maliyetlerin toplamı adımı, farklı şekillere ve ağırlık değerlerine sahip bir destek penceresi içindeki pikseller kullanılarak eşleşme maliyetleri elde edilmektedir. Destek bölgelerinin oluşturulması ile ilgili birçok yöntem bulunmaktadır. Örnek olarak bu yöntemlerden bazıları: kare pencere (square window) (Scharstein ve Szeliski, 2002), çoklu pencere (multiple window) (Adhyapak vd., 2007; Buades ve Facciolo, 2015), uyarlanabilir pencere (adaptive window) (Kanade ve Okutomi, 1991; Zhou vd., 2006; Shi vd., 2016) ve değiştirilebilir penceredir (shiftable

window) (Wang vd., 2006; Gong vd., 2007).

Yoon ve Kweon (2006), önerdiği uyarlanabilir destek ağırlığı (Adaptive Support-Weight, ASW) yönteminde destek penceresi içerisindeki bütün piksellerin ağırlıklarının aynı olmaması gerektiğini vurgulamıştır. Bu sebepten dolayı, Yoon ve Kweon renkler arasındaki benzerlik ve pencere içerisindeki pikseller ile merkez piksel arasındaki geometrik mesafeyi baz alarak, pencere içerisindeki piksellerin her birine belirli ağırlık değerleri atamıştır. Bu ağırlık değerlerinin belirlenmesinde ise bilateral filtre (Bilateral Filter, BF) (Tomasi ve Manduchi, 1998) yapısı kullanmıştır. Rhemann vd. (2011) ise çalışmasında yüksek kaliteli sonuçlar üreten ve bilateral filtreden daha iyi performans gösteren klavuzlu filtre (Guided Filter, GF) (He vd., 2013) tabanlı yeni bir algoritma önermiştir. Önerdikleri yapı ile işlem yükünü ve çalışma süresini kısaltırken ayrıca bilateral filtreli yapılara göre daha başarılı sonuçlar elde etmişlerdir.

Zhang vd. (2009), tarafından her şekle uyarlanabilir çapraz tabanlı maliyet toplama (Cross Based Cost Aggregation, CBCA) yöntemi önerilmiştir. CBCA yöntemi her pikselin yalnızca bulunduğu nesnenin sınırları içinde uyarlanabilir bir pencere alanı belirlenmesini sağlamaktadır. Bu uyarlanabilir pencerenin sınırları belirlenirken her bir pikselin konumunun çevresinde benzer görüntü yoğunluk değerlerine sahip piksellerden oluşan bir yerel alan seçilmektedir. Uyarlanabilir pencerelerin sınırları birbirini sıralı olarak takip eden iki adımda belirlenmektedir. İlk adımda, her piksel için dikey eksen de piksel komşuluklarındaki yoğunluk değişimlerine göre yerel destek bölgesinin dikey sınırları tanımlanmaktadır. İkinci adımda ise belirlenen dikey eksen de her bir piksel için yatay eksen de komşu pikseller arasındaki yoğunluk farklılıkları dikkate alınarak uyarlanabilir pencerenin sınırları belirlenmektedir.

Yerel yöntemler de eşitsizlik değerleri, eşleşme maliyetinden en küçüğü seçilerek elde edilmektedir ve bu yöntem “kazanan-hepsini-alır (Winner-Takes-All, WTA)” olarak adlandırılmaktadır. Yerel yöntemlerde WTA ile küresel yöntemlerde ise eniyileme yöntemleri aracılığıyla başlangıç eşitsizlik haritası elde edilmektedir. Bu başlangıç eşitsizlik haritası birçok yanlış eşitsizlik değeri (aykırı değer) içermektedir. Aykırı değer tespit etme ve eşitsizlik haritasını iyileştirme, stereo eşleştirme yöntemlerinde doğruluğu daha da artırmak için ihmal edilemez bir adımdır (Jiao vd., 2014; Zhan vd., 2015; Zhu ve Dai, 2017; Yan vd., 2019). Stereo eşleştirmede aykırı değerler, çoğunlukla

perdelemelerden (occlusions) veya yanlış eşleştirmelerden (mismatches) kaynaklanmaktadır. Eşleştirme doğruluğunu arttırmak için, çoğu yaklaşım eşitsizlik haritasını düzeltme adımı içermektedir (Jiao vd., 2014; Zhan vd., 2015). Fakat düzeltme adımından önce yanlış eşleşmiş (aykırı) noktaların tespiti de düzeltme işlemi kadar önemlidir (Hayashi vd., 2005; Dong ve Feng, 2019). Aykırı noktaların tespiti için sağ-sol tutarlılık (Left-Right Checking, LRC) (Scharstein ve Szeliski, 2002) testi sıklıkla kullanılan yöntemlerden bir tanesidir (Mattoccia vd., 2007). LRC yönteminde sol ve sağ görüntüyü ayrı ayrı referans görüntü olarak kabul ederek elde edilen iki farklı eşitsizlik haritasına ihtiyaç vardır.

Eşitsizlik haritası üzerindeki aykırı değerler tespit edildikten sonra bu değerlerin düzeltilmesi gerekmektedir. Bu amaç kapsamında literatürde farklı yöntemler kullanılmıştır. Peng vd. (2015), süperpiksel tabanlı bölütleme tekniği kullanan bir yöntem önermiştir. Bu yöntemde, eşitsizlik haritasında renk tabanlı bir kümeleme yapılmış ve bu kümeler içerisindeki aykırı noktalar düzeltilmiştir. Wang vd. (2006) tarafından aykırı değerleri azaltmak için temel Medyan Filtresi (Median Filter, MF) kullanılmıştır. Bunun yanında ağırlıklı medyan filtresi (Weighted Median Filter, WMF)'de aykırı değerleri düzeltmek için kullanılan bir diğer yöntemdir (He vd., 2013; Ma vd., 2013; Rhemann vd., 2011).

1.1.2 Derin öğrenme tabanlı yöntemler

Makine öğrenmesi uzun yıllardan beri sadece görüntü işleme ve makine görmesinde değil birçok alanda yoğun bir şekilde çalışılan bir alandır. Son zamanlarda ise derin öğrenme kavramı görüntü işleme alanında yaygın olarak çalışılan ve kullanılan bir yöntem olarak ön plana çıkmaktadır. Grafik işlem birimlerinin (Graphics Processing Unit, GPU) artan işlem kabiliyetleri sayesinde bu alandaki gelişmeler daha da hızlanmış durumdadır.

Derin öğrenme kavramı içerisinde pek çok teknik ve farklı yapıya sahip mimari barındırmaktadır. En basit hali ile derin mimariler çok sayıda katman içeren ve çoğunlukla doğrusal olmayan (non-linear) işlemler gerçekleştiren yapılar olarak tanımlanabilir. Evrimsel sinir ağı (ESA) (Convolutional Neural Network, CNN) adı verilen ağlar ise özellikle görüntü işleme alanında oldukça başarılı sonuçlar elde edebildiği için sıklıkla tercih edilmektedir. CNN'ler görüntü işlemede; görüntü

renklendirme (R. Zhang vd., 2016; R. Zhang vd., 2017), nesne algılama (Girshick vd., 2014; Ren vd., 2017) , yüz tanımlama (Y. Sun vd., 2014; Taigman vd., 2014), görüntü bölütleme (Koh ve Kim, 2017; Long vd., 2015), kenar belirleme (Xie ve Tu, 2017), süper çözünürlük (Dong vd., 2014; Dong vd., 2015; Ledig vd., 2017), ve stereo eşleştirme (Chang ve Chen, 2018; S. Kim vd., 2019; Mayer vd., 2016; Žbontar ve Lecun, 2015) gibi birçok konuda yoğun olarak kullanılmaktadır.

Makine öğrenmesi tabanlı yaklaşımlar derin öğrenme öncesinde de eşitsizlik haritalarının elde edilmesi ve düzeltilmesi işlemlerinde kullanılmıştır (Kong ve Tao 2004; Li ve Huttenlocher 2008; Scharstein ve Pal 2007). Bu çalışmalar genellikle eşleşme maliyeti tahmini veya enerji fonksiyonu parametrelerinin eniyilenmesini amaçlamıştır.

Son zamanlarda literatürdeki stereo eşleştirme yöntemleri incelendiğinde CNN tabanlı birçok farklı yaklaşım ön plana çıkmaktadır. Bu yöntemler genellikle eşleşme maliyetlerinin hesaplanması (Žbontar ve Lecun, 2015; Z. Chen vd., 2015; Luo vd., 2016; Kim vd., 2018; Brandao vd., 2019), aykırı noktaların tespit edilmesi ve eşitsizlik haritalarının düzeltilmesi (Poggi ve Mattoccia, 2016; Seki ve Pollefeys 2016; Poggi ve Mattoccia, 2017; Cheng vd. 2018; Kim vd. 2019; Kim vd., 2019; Gul vd., 2019) veya uçtan-uca öğrenme (end-to-end learning) ile eşitsizlik haritalarının elde edilmesi (Mayer vd., 2016; Pang vd., 2017; Kendall vd., 2017; Chang ve Chen 2018; Liang vd., 2018; Zhang vd., 2019; Kang vd., 2019; Nguyen and Jeon , 2019; Guo vd., 2019; Zhang vd., 2020) şeklinde olmuştur.

Uçtan-uca öğrenmeye dayalı yöntemlerde, önerilen ağlar giriş olarak aldıkları stereo görüntüler üzerinde bir takım işlemler gerçekleştirerek ağın çıkışında kullanıcılara doğrudan eşitsizlik haritası üretmektedir. Diğer yöntemlerde ise stereo eşleştirme işleminde yer alan adımlardan bir veya birkaçını derin ağlar yardımı ile gerçekleştirmek amaçlanmaktadır.

Literatürde yeralan çalışmalar; stereo eşleştirme işlemindeki adımlar için önerilen yöntemler “CNN tabanlı eşleşme maliyeti hesaplama” , “CNN tabanlı aykırı değer tespiti ve fark haritası düzeltme” ve uçtan uca ağ yapıları “CNN tabanlı uçtan-uca öğrenme” başlıkları altında incelenmiştir.

1.1.2.1 CNN tabanlı eşleşme maliyeti hesaplama

Stereo eşleştirme alanında ilk CNN tabanlı yöntem Žbontar ve Lecun (2015), tarafından eşleşme maliyetlerinin elde edilmesi amacıyla önerilmiştir. Önerilen yöntemde eşleşme maliyetleri derin siyam ağlarla (deep siamese network) hesaplanmıştır ve ağ yapısı MC-CNN olarak adlandırılmıştır. Sağ ve sol görüntüden alınan görüntü parçalarında siyam ağlar yardımı ile benzerlik karşılaştırmaları yapılarak ağ parametreleri eğitilmiştir. Görüntü parçaları, veri kümelerine bağlı olarak 9×9 veya 11×11 olarak seçilmiştir. Žbontar ve Lecun (2015) çalışmasında “Hızlı mimari” ve “Doğru mimari” olmak üzere iki farklı ağ mimarisi sunmuştur. Önerilen ağ mimarileri sadece maliyet hesabı yaptığı için eşitsizlik haritası elde etme işleminde geleneksel stereo eşleştirme algoritmalarından faydalanılmıştır. Maliyetlerin toplamı işlemi için çapraz tabanlı maliyet toplama (Cross Based Cost Aggregation, CBCA), eşitsizlik haritası elde etmek için ise yarı küresel eşleştirme (Semi Global Matching, SGM) algoritmaları kullanılmıştır. Nihai eşitsizlik haritası ise elde edilen eşitsizlik haritaları kullanılarak çoklu adımlardan oluşan bir düzeltme işlemi uygulanarak elde edilmektedir.

Z. Chen vd. (2015), Žbontar ve Lecun (2015)’nun çalışmasından ilham alarak eşleşme maliyetleri için benzerlik puanı hesaplama yöntemi önermiştir. Bu yöntemi MC-CNN’den ayıran iki temel fark vardır. Bunlardan ilki, önerilen ağ yapısı MC-CNN’den farklı olarak iki siyam ağ içermektedir. Bu ağlar giriş olarak 13×13 ve 7×7 ’lik iki farklı boyutta görüntü parçalarını giriş olarak almaktadır. İki siyam ağdan da karşılaştırma işlemlerinin sonunda öznitelik vektörleri elde edilmektedir. İkinci fark ise, CNN katmanlarının sonunda elde edilen öznitelik vektörlerinden (feature vector) benzerlik puanı noktasal çarpımla doğrudan elde edilmektedir. İki farklı boyuttan elde edilen karşılaştırmalar ise $1 \times 1 \times 2$ ’lik bir evrişim işlemi ile birleştirilmiştir. Ağın eğitimi için ise regresyon modeli uygulanmıştır ve öklid (euclidean) maliyeti kullanılmıştır.

Luo vd. (2016), GPU üzerinde hesaplaması bir saniyeden daha kısa süren ve oldukça doğru sonuçlar üretebilen bir eşleştirme ağı önermişlerdir. Bu ağ mimarisinde giriş görüntülerine siyam ağlarda sıralı evrişim işlemleri uygulanarak öznitelikler elde edilmektedir. Elde edilen özniteliklerin benzerliklerinin hesaplanması için korelasyon (ilinti) işleminden faydalanılmıştır. Ayrıca ağ mimarisi, olası her eşitsizlik değerleri için ayrı ayrı sınıflandırma yapabilecek şekilde eğitilmiştir. Ağın girişlerinin birine sol

görüntüden alınan pencere verilirken, diğer girişe ise sağ görüntüden arama uzayı boyunca bütün muhtemel olasılığı içeren görüntü kümesi verilmiştir. Böylece, bütün olası eşitsizlik değerleri için bir olasılık değeri hesaplamışlardır. Diğer mimarilerden farklı olarak çıkışta tek bir eşleştirme için benzerlik sonucu elde etmek yerine bütün muhtemel sonuçlar için eşleştirme sonucu üretmişlerdir. Eşitsizlik haritası elde edilmesi işleminde ise Žbontar ve Lecun'nun (2015) gerçekleştirdiği gibi geleneksel yöntemlerden faydalanmışlardır.

Kim vd. (2018), stereo eşleştirme için CNN tabanlı dört farklı öznitelik çıkarımı için dört farklı ağ mimarisi önermiştir. Çıkarılan özniteliklerden ikisi, farklı boyutlara sahip çok ölçekli özniteliklerdir. Bu çok ölçekli öznitelikler eşleşme maliyetlerini hesaplamak için kullanılmaktadır. Ardından çok ölçekli öznitelikler ve kenar bilgisi birlikte kullanılarak en uygun maliyetin belirlenmesi amaçlanmaktadır.

Brandao vd. (2019) tarafından sunulan çalışmada, (Luo vd. 2016) tarafından geliştirilen çalışmaya çok benzer bir ağ mimarisi önerilmiştir. Fakat bu çalışmadan iki büyük farkı vardır. Bunlardan ilki, korelasyon işleminden önce uygulanan havuzlama işlemi esnasında kaybolan detaylar ters-evrişim işlemi ile düzeltilmiştir. İkinci fark ise, korelasyon öğrenme problemini büyük ölçüde basitleştiren bir öznitelik birleştirme yöntemi ile çıkarılan özniteliklerin kalitesinin daha iyi değerlendirilmesine olanak sağlanmıştır.

1.1.2.2 CNN tabanlı aykırı değer tespiti ve fark haritası düzeltme

Literatürde CNN tabanlı aykırı değer tespiti ve fark haritası düzeltme alanında farklı çalışmalar mevcuttur. Bu çalışmalardan, Poggi ve Mattoccia (2016) sadece tek bir eşitsizlik haritası üzerinden değerlendirme yaparak güven değeri elde etmiştir. Önerdikleri ağ mimarisi giriş olarak eşitsizlik haritası üzerinden 9×9 'luk bir görüntü parçası almaktadır. Giriş görüntüsüne uygulanan evrişim işlemi sonunda ise tam bağlı (Fully Connected - FC) katmanlar kullanılarak eşitsizlik değeri belirlenmiştir. Ağın yapısındaki çıkış katmanına sigmoid fonksiyonu uygulanarak güven değerleri 0 – 1 aralığında ölçeklendirilmiştir.

Seki ve Pollefeys (2016) tarafından önerilen yöntemde, geleneksel yöntemlerle eşitsizlik haritası elde edilmesi ile derin ağlarla aykırı değerlerin tespit edilmesi fikirleri birleştirilerek tasarlanmıştır. Önerdikleri yöntemde eşitsizlik haritalarının elde edilmesinde kullanıldıkları yarı küresel eşleştirme yönteminin ayarlanabilir parametrelerini derin ağlarla elde ettikleri güven değerleri ile ilişkilendirmişlerdir. Eşitsizlik haritası elde edebilmek için güven değerleri doğrudan yarı küresel eşleştirme yönteminin ayarlanabilir parametrelerine dâhil edilmiştir. Ağın girişinde 3B düzlemde aynı noktayı temsil eden sol ve sağ eşitsizlik haritalarından alınan iki kanallı görüntü parçaları kullanılmaktadır. Bu görüntüler merkez piksel değeri çerçevesinde ön işleme tabi tutulmaktadır. Daha sonra, sıralı evrişim işlemleri uygulanarak tutarlılık değeri elde edilmektedir.

Poggi ve Mattoccia (2017) ise, 2016 yılındaki yaptıkları çalışmaya (Poggi ve Mattoccia 2016) benzer bir ağ mimarisi önermişlerdir. Önerdikleri diğer yapıdan farklı olarak evrişim işlemlerinin sonunda bulunan tam bağlı katmanlar yerine tekrar evrişim işlemleri ile devam edilerek güven değeri hesaplanmıştır.

Cheng vd. (2018) diğer önerilen yapılardan farklı olarak doğrudan eşitsizlik haritası ve referans giriş görüntülerine göre düzeltme işlemi yapmak yerine maliyet yığınlarını da dikkate alarak yanlış eşleşen pikselleri tespit etmeye çalışmışlardır. Bu mimariyi oluştururken, Cheng vd. (2018)'nin temel amaçları yanlış eşleşen noktaları tespit etmek ve yanlış eşleşen noktaları doğru eşitsizlik değerleri ile düzeltmektir. Ayrıca, maliyet yığınları ve başlangıç eşitsizlik haritasının elde edilmesi işlemi için Mayer vd. (2015) tarafından önerilen DispNet'den faydalanılmıştır. Bu ağ yapısı ConfNet (güven ağı) ve GradNet (eğim ağı) olmak üzere harici iki alt-ağ içermektedir.

ConfNet (güven ağı), referans görüntüsünü, DispNet ağından elde edilen maliyet yığınlarını ve LRC'den elde edilen bilgileri birleştirerek giriş olarak almaktadır. ConfNet yapısında her piksel için çıkış katmanı, iki çıktılı softmax fonksiyonudur. Bu ağın eğitimi için çapraz entropi (Cross Entropy, CE). ile ağın kayıp değerleri belirlenmiş ve Adam (Kingma ve Ba, 2014) temelli rasgele gradyan inişi (Stochastic Gradient Descent, SGD) algoritması yardımı ile eniyileme gerçekleştirilmiştir. GradNet ise giriş olarak referans görüntüsünü ve DispNet ağından elde edilen maliyet yığınlarını almaktadır. GradNet'in çıkış katmanında ise her bir piksel için iki nöron bulunmaktadır. Bu nöronlar ilgili piksel

için yatay ve dikey eksenlerdeki gradyan değerlerini göstermektedir. Giriş görüntüsüne ait doğru gradyan değerleri olmadığı için, bu değerler doğru eşitsizlik haritası üzerinden yatay ve dikey yönde piksel farkları göz önünde bulundurularak oluşturulmuştur. L1 (Manhattan distance, Manhattan mesafesi) kayıp fonksiyonu kullanılarak Adam (Kingma ve Ba, 2014) yöntemi ile GradNet ağ yapısının parametreleri eğitilmiştir. Son adımda nihai eşitsizlik haritası elde edilmesinde küresel yöntemlerden faydalanılmıştır. Küresel yöntemde kullanılan markov rastgele alanı (Markov Random Field, MRF) yapısı başlangıç değerlerini iki ağdan elde edilen verilerden ve başlangıç eşitsizlik haritasından sağlamıştır.

Kim vd. (2019) ise çalışmalarında Cheng vd. (2018) gibi güven değeri elde etmek için sadece eşitsizlik haritası kullanmanın başarılı sonuçlar elde etmek için yeterli olmadığını belirtmiştir. Bundan dolayı önerdikleri ağ mimarisinde doğrudan eşitsizlik haritası kullanmak yerine ham eşleşme maliyeti kullanmışlardır. Ham eşleşme maliyeti ise MC-CNN'den (Žbontar ve LeCun 2015) elde edilmiştir.

LAF-Net (Kim vd., 2019). ağında ise Cheng vd. (2018)'nin çalışmasında olduğu gibi aykırı noktaların tespit edilmesinde sadece eşitsizlik haritaları kullanılmamıştır. Önerdikleri ağ yapılarında giriş olarak eşleşme maliyetleri, eşitsizlik haritası ve referans sol görüntü alınmıştır.

Confidence-RNN (C-RNN) ağ yapısında, uzun kısa süreli bellek (Long Short-Term Memory, LSTM) ağları kullanılarak eşleşme maliyet hacimlerindeki güven değerleri tahmin edilmiştir (Gul vd., 2019). Her bir piksel değeri için güven değeri, eşleşme maliyetlerinin LSTM ağında değerlendirilmesi ile verilmiştir.

Wang vd. (2022) stereo eşleştirme işlemindeki aykırı noktalarını Kiureghian ve Ditlevsen, (2009)'in çalışmasında ortaya attığı belirsizlikler olarak tanımlamıştır. Kiureghian ve Ditlevsen, (2009) çalışmasında belirsizliklerin kaynağının mevcut olmamasına rağmen bu belirsizlikleri genellikle tesadüfi (tesadüfi) veya epistemik (epistemic) olarak sınıflandırmıştır. Wang vd. (2022), hem tesadüfi hem de epistemik belirsizlikleri tahmin etmek için uçtan uca bir ağ yapısı önermiştir. Ayrıca belirsizlikleri hesaplamak için kullanılabilen Normal Ters-Gama (Normal Inverse-Gamma, NIG) dağılımı olarak adlandırdıkları bir dağılım sunmuşlardır.

1.1.2.3 CNN tabanlı uçtan-uca öğrenme

Uçtan-uca öğrenme ile eşitsizlik haritası üreten ağ mimarisi ilk olarak Mayer vd. (2016) tarafından önerilmiştir. Bu ağ mimarisi DispNet olarak adlandırılmıştır. Ayrıca, aynı çalışmada optik akış (optical-flow) ve stereo eşleştirme problemlerinde kullanılabilecek bir sentetik veri kümesi de önerilmiştir. Önerdikleri ağ mimarisi kodlayıcı-kod çözücü mimarisini içermektedir ve ağ yapısını önerirken FlowNet (Dosovitskiy vd., 2015) mimarisinden ilham almışlardır. Kod çözücü katmanlarının her birinden bir adet alt ölçekli eşitsizlik haritası elde edilmiştir. Eğitim esnasında her bir eşitsizlik haritasından elde edilen hata bilgisine göre ağın eğitilmesi zor olacağı için eğitime ilk olarak en küçük eşitsizlik haritasından başlanmıştır. Böylelikle ağın önce kaba temsili öğrenmesi amaçlanmıştır. Ardından sıra ile diğer eşitsizlik haritaları kullanılarak ağın daha ince ayrıntıları öğrenmesi sağlanmıştır. Ağın eğitimi için ise Adam (Kingma ve Ba, 2015) yöntemi kullanılmıştır.

Pang vd. (2017) ise Mayer vd. (2016) tarafından yapılan çalışmayı geliştirip, iki aşamadan oluşan CRL (Cascade Residual Learning) ağ mimarisini önermiştir. Bu mimaride her iki aşamada da eşitsizlik haritası elde edilmiştir. İlk aşamada oluşturulan ağ yapısı, Mayer vd. (2016) tarafından önerilen DispNet ağ mimarisine ek bir ters evrişim modülü eklenerek elde edilmiştir ve bu mimari DispFulNet olarak adlandırılmıştır. İkinci aşamada ise DispResNet ağı bulunmaktadır ve bu ağ birinci aşamada üretilen eşitsizlik haritasının düzeltilmesi için kullanılmaktadır. Bu düzeltme işleminde kum saati (hourglass) ağ mimarisi kullanılarak çok ölçekli bir düzeltme işlemi yapılmaktadır. Çıkışta nihai eşitsizlik haritası bu iki ağdan elde edilen eşitsizlik haritasının birleştirilmesi ile elde edilmiştir.

Başka bir ağ mimarisi Kendall vd. (2017) tarafından sunulmuş olup, GC-Net (Geometry and Context Network) mimarisi olarak adlandırılmıştır. GC-Net yapısında 3B evrişimler kullanılarak maliyet yığının düzenlenmesi için uçtan-uca bir ağ yapısı önerilmiştir. Eşitsizlik değerleri ise, maliyet yığını üzerinden önerdikleri türevi alınabilir soft-argmin yöntemi ile elde edilmiştir. Böylece, herhangi bir ek işlem veya düzeltme işlemi olmadan alt piksel (sub-piksel) doğruluğuna kadar uçtan-uca eğitilebilir bir yöntem elde edilmiştir.

Chang ve Chen (2018), içerik bilgisinin iyi tanımlanamadığı bölgelerde eşitsizlik

bilgisinin doğru bulunamadığını belirtmiştir. Bu nedenle, içerik bilgisinin daha iyi algılanarak başarımın iyileştirilmesi için piramit stereo eşleştirme ağı (PSMNet, pyramid stereo matching network) önerilmiştir (Chang ve Chen, 2018). Bu ağ mimarisi, uzamsal piramit havuzu (Spatial Pyramid Pooling, SPP) ve 3B CNN olmak üzere iki parçadan oluşmaktadır. Uzamsal piramit havuzu sayesinde farklı boyutlardan ve konumlardan küresel içerik bilgisi çıkarılarak maliyet yığınları oluşturulmuştur. 3B CNN ise eşleşme maliyetlerinden yığın kum saati (stacked hourglass) yapısını kullanarak eşitsizlik haritaları elde etmek için kullanılmıştır.

Liang vd. (2018) tarafından önerilen yapıda ise Scharstein ve Szeliski (2002) tarafından önerilen stereo eşleştirme adımları birleştirilerek ağ mimarisi oluşturulmuştur. Önerilen ağ mimarisi dört parçadan oluşmaktadır. İlk stereo eşleştirme adımında, çok ölçekli paylaşılan ağırlıklar kullanılarak öznitelikler elde edilmektedir. İkinci adımda, elde edilen öznitelikler kullanılarak başlangıç eşitsizlik haritası oluşturulmaktadır. Üçüncü adımda, başlangıç eşitsizlik haritası ve öznitelikler kullanılarak her bir piksel için tutarlılık (consistency) değerleri elde edilmektedir. Son adımda ise, tutarlılıklar ve başlangıç eşitsizlik haritası kullanılarak nihai eşitsizlik haritası üretilmektedir.

Eşitsizlik haritalarında nesnelerin geçiş bölgelerini daha keskin hale getirebilmek için kenar ve bölütleme bilgileri de kullanılmıştır. Yang vd. (2018) SegStereo ağında kayıp fonksiyonuna bölütleme ipuçları ekleyerek, Song vd. (2018) EdgeStereo ağında kayıp fonksiyonuna kenar ipuçları ekleyerek eşitsizlik haritasının geçiş bölgelerindeki eşitsizlik değerlerini düzeltmeyi amaçlamıştır.

Zhang vd. (2019), GANet ağ yapısı ile stereo eşleştirme problemlerinde sıklıkla kullanılan yarı küresel eşleştirme yöntemine türevlenebilir bir yaklaşım getirerek uçtan uca bir ağ yapısı geliştirmiştir. Ayrıca çalışmalarında ince yapıları iyileştirmek için geleneksel bir maliyet filtreleme stratejisini izleyen yerel güdümlü toplama (Local Guided Aggregation, LGA) katmanını önermiştir.

Zhang vd. (2019) eşitsizlik haritasındaki her bir pikselin güvenilirliğini puanlandırarak güven haritası elde etmiştir. Güven haritasını elde ederken PSMNet'den elde edilen maliyet hacimleri ile önerdikleri güven tahmin ağı (CENet) kullanılmıştır. Daha sonra gerçek değerlerinin bilinmesi mümkün olmayan doğru eşleşme maliyet hacmi elde etmek

amacıyla güven değerleri ve doğru eşitsizlik değerleri kullanarak tahmini bir eşleşme maliyet hacmi elde edilmiştir. Zhang vd. (2019) önerdikleri AcfNet ağ yapısını eğitirken üç adet kayıp fonksiyonun birleşimini almaktadır. Bunlar kestirilen doğru maliyet hacimlerinden elde edilen kayıplar, maliyet hacimleri üzerinden hesapladıkları güven değerlerinden elde edilen kayıp değerleri ve PSMNet ağından üretilen eşitsizlik haritasından elde edilen kayıp değerleridir.

Kang vd. (2019) ise, maliyet hacmi oluştururken küresel içerik bilgisini daha iyi bir şekilde ifade edebilmek için farklı boyutlarda genişletilmiş (dilated) evrişim işlemleri kullanmıştır. Daha sonra, korelasyon işlemleri ile elde edilen maliyet hacimlerinin üzerinde kodlayıcı-kod çözücü yapısını kullanarak farklı boyutlarda eşitsizlik haritaları elde edilmektedir. DispNet (Mayer vd. 2015) mimarisinde olduğu gibi kod çözücü yapısı ile her bir katmanda kayıp ifadeleri üreterek eğitimi kolaylaştırmayı ve iyileştirmeyi amaçlamışlardır. Nihai kayıp fonksiyonunu ise, L1 kayıpları ile gradyan kayıplarını belirli bir kaysayı oranı ile birleştirilmesiyle elde edilmiştir.

Nguyen and Jeon (2019) iki aşamadan oluşan bir ağ yapısı önermiştir. Önerdikleri ağ yapısı ilk aşamada geniş çapta öznitelikleri öğrenmeyi amaçlamaktadır. Geniş çaplı öznitelik çıkarma aşamasında PSMNet’de (Chang and Chen,2018) kullanılan uzaysal havuzlama yöntemi ile beraber genişletilmiş (dilated) evrişim yapısı kullanılmıştır. Bu işlemin amacı ise, küresel içerik bilgisinin en iyi şekilde elde edilmesi olarak ifade edilmiştir. İkinci aşamada ise bu öznitelikler kullanılarak decoder-encoder yapısı ile eşitsizlik haritası elde edilmiştir.

Büyük ağ mimarileri daha iyi temsil yetenekleri ile daha iyi sonuçlar üretme konusunda ne kadar başarılı olsalar da bu yöntemlerin işlem maliyetleri onları gerçek zamanlı uygulamalar için verimsiz kılmaktadır. Bunun üstesinden gelmek için Duggal vd. (2019) önerdikleri DeepPruner ağında, arama alanını türevlenebilir bir PatchMatch katmanıyla budayarak eşleştirme maliyetinin boyutlarını düşürmüştür. Bu sayede PSMNet ve GC-Net gibi 4B maliyet hacimlerinin kullanıldığı yapılarda neredeyse on kat küçük bir maliyet hacminin elde edilmesi sağlanmıştır.

Başka bir çalışmada önerilen GwcNet ağı ile korelasyon işlemi yapılırken maliyet hacimleri doğrudan korelasyon işlemine tabi tutulmak yerine gruplar halinde korelasyon

yapılarak DispNet ağıının performansı artırılmıştır (Guo vd., 2019). Ayrıca AANet’de farklı çözünürlüklere sahip öznetelik haritalarında korelasyon işlemi gerçekleştirilerek üç farklı maliyet hacmi oluşturmaktadır (Xu and Zhang,2020). Daha sonra bu maliyet hacimleri verimli ve doğru stereo eşleştirme için önerdikleri seyrek noktalara dayalı ölçek içi maliyet toplama (Intra-Scale Cost Aggregation, ISA) modülü ve çapraz-ölçek maliyet toplama (Cross-Scale Cost Aggregation, CSA) modüllerine aktarılmıştır.

Bir başka çalışmada ise, stereo eşleştirme ağ yapılarının daha önce karşılaşmadıkları veri kümelerinde veya renk, aydınlatma, kontrast ve doku gibi farklılıkların oluşması durumunda başarılarının düştüğü ortaya konmuştur. Zhang vd. (2020) önerdikleri DSM-Net ağ yapısıyla alan normalleştirme ile farklı görüntü şartlarında veya farklı ortamlardaki görüntülerde ağın daha başarılı sonuçlar elde etmesini sağlamışlardır.

Cheng vd. (2020) LEAStereo ağında sinir mimarisi arama (Neural architecture search, NAS) teknikleri kullanılarak uçtan uca bir stereo eşleştirme yapısı önermiştir. Önerdikleri yapıda öznetelik çıkarma ve eşleme maliyetlerini elde etmek amacıyla sinir mimarisi arama (Neural Architecture Search, NAS) tabanlı FeatureNet ve MatchingNet yapılarını önermişlerdir.

Chen vd. (2021) önerdikleri PGNet ağında panoptik bölütleme bilgilerini kullanarak eşitsizlik haritası elde etmektedirler. Panoptik bölütleme anlamsal bölütleme ve örnek bölütlemenin birleşimi olarak bilinmektedir. PGNet ağında PSMNet ağı omurga olarak seçilmiş ve PSMNet ağına güven, artık ve kayıp olmak üzere üç modül eklenmiştir.

Yukarıda verilen üç ana başlık dışında da derin ağlar kullanılarak eşitsizlik haritalarının elde edilmesi ve iyileştirilmesi için yapılan çalışmalar mevcuttur. Bunlardan biri Seki ve Pollefeys, (2017) tarafından önerilen SGM-Net mimarisidir. Bu mimari SGM algoritmasının parametrelerini giriş görüntülerine göre adaptif bir şekilde değiştirmeyi amaçlamaktadır. Williem ve Park, (2018) ise, census dönüşümü ve Hamming mesafesi (Hamming distance) ile elde ettikleri eşleşme maliyetlerini filtrelemek için derin ağ mimarisi önermiştir. Bu yöntem geleneksel yöntemlerin aksine, filtrenin ağırlık değerlerini belirlemek için renkli giriş görüntüsüne ihtiyaç duymadan doğrudan maliyet yığınları üzerinden filtreleme işlemi yapmaktadır.

1.2 Tez Kapsamında Gerçekleştirilen Çalışmalar

Bu tez çalışması kapsamında stereo görüntüler kullanılarak eşitsizlik haritalarının elde edilmesi ve eşitsizlik haritalarının düzeltilmesi amacıyla farklı yöntemler önerilmiştir (Emlek ve Peker, 2020; Emlek ve Peker, 2021a; Emlek ve Peker, 2021b). Ayrıca, önerilen çalışmalarda özellikle gerçek zamanlı uygulamalar ve otonom sistemlerin gereksinimleri göz önünde bulundurulmuştur. Dolayısıyla önerilen yapıların çalışma süreleri gerçek zamanlı uygulamalar için mümkün olduğunca düşük tutulmaya çalışılmıştır. Ayrıca otonom sistemlerin çoğu mobil bir platform tarafından kontrol edildiği düşünülerek, önerilen yapıların parametre sayısı ve GPU üzerinde katladıkları alanların çok büyük boyutta olmamasına dikkat edilmiştir.

Bölüm 2’de literatürdeki geleneksel ve derin öğrenme tabanlı çalışmalar incelenerek öne çıkan eksiklikler değerlendirilmiş ve çalışmalarda önerilen ağların eğitim ve test işlemlerinde kullanılan veri kümeleri tanıtılmıştır. Ayrıca, bu bölümde önerilen ağların literatürdeki diğer yöntemlerle karşılaştırmasında kullanılan yöntemler ve deney sistemi hakkında da bilgi verilmiştir.

Bölüm 3’te (Emlek ve Peker, 2020), eşleşme maliyetlerine erişmenin mümkün olmadığı kapalı veya hazır sistemlerden elde edilen eşitsizlik haritalarının düzeltilmesi amacıyla bir ağ yapısı önerilmiştir. Önerilen bu yapıda eşitsizlik haritalarının düzeltilmesinde sadece stereo görüntüler ve eşitsizlik haritaları kullanılmıştır.

Bölüm 4’te (Emlek ve Peker, 2021a), stereo eşleştirme işleminde stereo görüntülerin aynı yatay ekseninde bulunmasından dolayı eşleşme maliyetlerinin paylaştığı bu sıralı bilgiler kullanılarak eşitsizlik haritası elde edilirken oluşan yanlış eşitsizlik değerlerinin düzeltilmesini mümkün kılacak bir ağ yapısı önerilmiştir. Bu ağ yapısıyla piksellerin ve dolayısıyla eşleşme maliyetlerinin sıralı yapısı kullanılarak eşleşme maliyetlerinin düzeltilmesi amaçlanmıştır.

Bölüm 5’te (Emlek ve Peker, 2021b) yatay eksenindeki piksellerin içerdiği sıralı bilgilerin stereo eşleştirme işlemlerinde ki öznitelik çıkarma katmanları için ekstra bilgi sağlaması amacıyla bir evrişim katmanı yapısı önerilmiştir.

Bölüm 6’da stereo görüntülerdeki bölgesel ve küresel özniteliklerin daha iyi çıkarabilmesi ve birleştirebilmesi amacıyla geleneksel öznitelik çıkarma yöntemlerinden esinlenerek bir modül önerilmiştir. Ayrıca önerilen modül kullanılarak gerçek zamanlı uygulamalar için uçtan uca eşitsizlik haritası elde edilmesi amacıyla bir ağ yapısı önerilmiştir.

Bu tez çalışması kapsamında yapılan çalışmalar ile eşitsizlik haritalarının ve eşleşme maliyetlerinin düzeltilmesini sağlayan ağ yapıları önerilmiştir. Bu ağ yapılarının yanı sıra stereo eşleştirme problemlerinde kullanılan derin öğrenme tabanlı öznitelik çıkarma işlemlerinin problem bazında özelleştirilmesini sağlayan bir yapıda önerilmiştir. Ayrıca gerçek zamanlı çalışmalar için yüksek doğrulukta sonuçlar üretebilen uçtan uca bir ağ yapısında önerilerek literatüre katkı sağlanmıştır.

BÖLÜM II

PROBLEM, VERİ VE SİMÜLASYON

Bu bölüm kapsamında ilk olarak literatürde yer alan stereo eşleştirme işlemlerindeki geleneksel ve derin öğrenme tabanlı çalışmalarda gözlemlenen eksiklikler verilmiştir. Bu gözlemler “Genel Değerlendirmeler” başlığı altında sıralı olarak sunulmuştur. Daha sonra “Veri Kümeleri” başlığı altında, tez çalışmasında önerilen ağların eğitim ve test işlemlerinde kullanılan veri kümelerinin tanıtımları yapılmıştır. “Performans Ölçütleri” başlığı kapsamında, literatürdeki yöntemler ve önerilen ağların başarımları değerlendirilirken kullanılan matematiksel yöntemler tanıtılmaktadır. “Nicel ve Nitel Karşılaştırmalar” başlığı altında, performans ölçütlerindeki çizelgelerin açıklamaları ve görsel karşılaştırmalarda kullanılan veri kümelerinin özel durumları verilmiştir. “Deney Sistemi” başlığı altında ise ağların eğitim ve test işlemlerinin gerçekleştirildiği donanım hakkında bilgi verilmiştir.

2.1 Genel Değerlendirmeler

Literatürdeki çalışmalar incelendiğinde, geleneksel ve derin öğrenme tabanlı yöntemlerin benzer konularda yanlış eşitsizlik değeri ürettikleri gözlemlenmiştir. Eşitsizlik haritalarındaki bu hatalı eşleşmelere sebep olan başlıca faktörler ve literatürde yer alan stereo eşleştirme alanındaki gelişmelerde gözlemlenen bazı negatif durumlar aşağıda sıralanmıştır;

- ❖ Geleneksel yöntemlerde olduğu gibi derin ağ tabanlı stereo eşleştirme yöntemlerinde de dokusuz alanlar ve perdelemelerin (occlusions) olduğu bölgelerde yanlış eşleşmeler daha çok ortaya çıkmaktadır.
- ❖ Ayrıca hareketli nesnelerin görüntülerinin elde edilmesi esnasında görüntülerde oluşan bulanıklaşmalardan dolayı o bölgelerdeki eşitsizlik bilgisi hesaplanırken hatalar oluşmaktadır.
- ❖ Ayna veya metal yüzeyler gibi ışığı yansıtan cisimler ve saydam cisimler görüntülerde ilgili noktalarda büyük yoğunluk farklılıklarına neden olarak önemli eşitsizlik hataları oluşturabilmektedir.
- ❖ Ağ modelleri, referans piksele karşılık gelecek piksel değerini arama işlemi esnasında, arama düzlemi boyunca tüm piksellerin eşitsizlik değerlerini veya

maliyetlerini işleme almamaktadır. Literatürdeki yöntemler incelendiğinde ise, tüm arama uzayındaki yatay ekseninde bulunan desen bilgisinin kullanılmadığı gözlenmektedir.

- ❖ Bazı ağ yapıları, farklı veri kümeleri için farklı parametreler ve yapılar kullanmaktadır. Bu durum ağların her ortamda başarılı sonuçlar üretmesini engellemektedir.
- ❖ Bazı ağ modellerinde ise, eşleşme maliyeti çıkartılma işlemleri sırasında sadece belirli bir pencere ile bölgesel bilgiler kullanılarak maliyetler elde edilirken küresel içerik bilgisinden faydalanılmamaktadır. Bu durum başarıyı kısmen düşürmektedir. Bu nedenle, son zamanlardaki çalışmalar öznitelik çıkarma işlemlerinde bölgesel ve küresel içerik bilgilerinin daha iyi temsil edilebilmesi üzerine yoğunlaşmıştır.
- ❖ Küresel içerik bilgisini aktarmak için son zamanlarda encoder-decoder yapısı, piramit havuzlama, genişletilmiş evrişim (dilated convolution) gibi yöntemler kullanılmıştır. Bu yöntemler belirli bir görüntü parçası yerine tüm görüntüyü kullanarak işlem yaptıkları için GPU üzerinde daha fazla yer kaplamaktadırlar.

2.2 Veri Kümeleri

Derin öğrenme ile birlikte büyüyen ağ yapılarının eğitilebilmesi için çok daha fazla veriye gereksinim oluşmuştur. Bununla birlikte veri kümelerindeki görüntülerin sayısı ile birlikte çeşitliliğinin miktarıda ağların başarımları üzerinde önemli etkiye sahiptir.

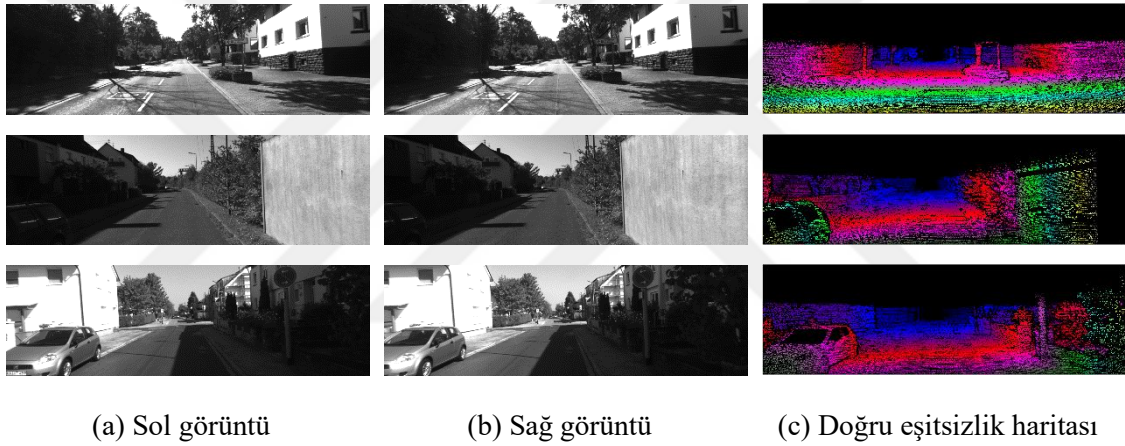
Çizelge 2.1’de stereo eşleştirme işlemlerinde eğitim ve test işlemleri sırasında en çok kullanılan veri kümeleri ve bu veri kümelerine dair bazı özellikler verilmiştir.

Çizelge 2.1. Literatürde en çok kullanılan veri kümeleri ve özellikleri

Veri kümeleri		Senaryo	Görüntü Sayısı		Eğitim sahneleri	Çözünürlük
			Eğitim	Test		
Middlebury (Scharstein vd., 2014)		İç mekân	23	10	23	2880 x 1988
KITTI	KITTI 2012 (Geiger vd., 2012)	Sürüş	194	195	194	1226 x 370
	KITTI 2015(Menze ve Geiger,2015)	Sürüş	200	200	200	1242 x 375
Scene Flow	FlyingThings3D (Mayer vd., 2016)	Sanal	21818	4248	2247	960 x 540
	Monkaa (Mayer vd., 2016)	Sanal	8591	-	8	960 x 540
	Driving (Mayer vd., 2016)	Sanal	4392	-	1	960 x 540
DrivingStereo (Yang vd., 2019)		Sürüş	174437	7751	38	1762 x 800

Bu tez çalışmasında Çizelge 2.1’de verilen veri kümeleri arasından, eğitim ve test işlemleri için Scene Flow (Mayer vd., 2016), KITTI 2012 (Geiger vd., 2012) ve KITTI 2015 (Menze ve Geiger, 2015) veri kümeleri kullanılmıştır.

KITTI 2012 ve KITTI 2015 veri kümeleri otonom sürüş uygulamalarında kullanılabilecek düzeltilmiş dış mekân görüntülerinden oluşmaktadır. Veri kümeleri sırasıyla eğitim için 194 ve 200, test için ise 195 ve 200 düzeltilmiş dış mekân görüntüsü içermektedir. Ayrıca, her iki veri kümesinde bulunan bütün stereo görüntü çiftleri için lazer tarayıcı ile toplanmış doğru derinlik değeri içeren eşitsizlik haritaları bulunmaktadır. KITTI 2012 ve KITTI 2015 veri kümelerine ait örnek stereo görüntüler ve doğru eşitsizlik haritaları Şekil 2.1 ve Şekil 2.2’de verilmiştir.



Şekil 2.1. KITTI 2012 veri kümesindeki eğitim kümesinden örnek görüntüler



Şekil 2.2. KITTI 2015 veri kümesindeki eğitim kümesinden örnek görüntüler

KITTI veri kümelerindeki eğitim görüntülerinin miktarı, büyük kapasiteli ağları eğitmek için oldukça küçük kalmaktadır. Dolayısıyla derin ağların eğitimi için Scene Flow veri

kümesi kullanılmıştır. Bu veri kümesi FlyingThings3D, Driving ve Monkaa olmak üzere üç alt veri kümesi içermektedir. Bu alt veri kümeleri sıra ile 21818, 8591 ve 4392 adet eğitim verisi içermektedir FlyingThings3D veri kümesi 4248 adet test görüntüsü içerirken Driving, Monkaa veri setleri test görüntüsü içermemektedir. Ayrıca bu veri setlerindeki tüm modüllerde yer alan eğitim ve test görüntülerinin çözünürlükleri 960×540 piksel boyutundadır. Scene Flow veri kümelerine ait örnek stereo görüntüler ve doğru eşitsizlik haritaları Şekil 2.3’de verilmektedir.

KITTI test veri kümelerinde yer alan doğru eşitsizlik haritaları kullanıcıların erişimine sunulmamaktadır. Test işlemleri web sitelerinde sunulan ara yüzde çevrimiçi olarak gerçekleştirilmektedir. Scene Flow veri kümesinde ise test veri kümelerinde bulunan görsellere ait doğru eşitsizlik haritaları verilmektedir.

2.3 Performans Ölçütleri

Tez kapsamında, farklı ağ yapıları ile elde edilen eşitsizlik haritaları nicel ve nitel karşılaştırmalar kullanılarak değerlendirilmiştir. Nicel karşılaştırmalar için, kötü eşleşen piksellerin yüzdesi (B) (Scharstein ve Szeliski, 2002), uç nokta hatası (End Point Error, EPE) (Otte ve Nagel, 1994) ve küresel fark (Global Difference, GD) (Yang vd., 2019) olmak üzere üç farklı değerlendirme metodu kullanılmıştır. B ve EPE, eşitsizlik haritasını değerlendirmek için kullanılan küresel hata ölçüleridir. B, belirlenen bir eşik değere göre oluşan hata sayısının toplam piksel sayısına oranını hesaplamaktadır. EPE, toplam hata içindeki büyük yerel hataların ağırlığını vurgular. Bir başka deyişle, EPE eşitsizlik bilgilerindeki hataları piksel mesafeleri cinsinden ölçerken, GD yöntemi ise 3B düzlemdeki derinlik değerlerinde meydana gelen hataları değerlendirmektedir. Bu tez kapsamında verilen çizelgelerde B değerlerinin 5,4,3,2 gibi farklı eşik değerleri için kötü eşleşen piksel hata oranları verilmektedir.

Uç nokta hatası Otte ve Nagel (1994) tarafından önerilmiş olup elde edilen eşitsizlik değerleri ile doğru eşitsizlik değerleri arasındaki mutlak farklar değerlendirilerek Denklem (2.1)’de verildiği gibi hesaplanmaktadır.



(a) Sol görüntü

(b) Sağ görüntü

(c) Doğru eşitsizlik haritası

Şekil 2.3. Scene Flow veri kümesinde yer alan eğitim kümesinden örnek görüntüler

$$E_{EPE_{disp}} = \frac{1}{N} \sum_{i \in \Omega} E_{disp}^i \quad (2.1)$$

Denklem (2.1)'de verilen N normalizasyon terimi olup geçerli tüm piksellerin (Ω) sayısını ifade etmektedir. E_{disp} ifadesi ise Denklem (2.2)'de verildiği gibi hesaplanmaktadır.

$$E_{disp} = |d_e - d_g| \quad (2.2)$$

Denklem (2.2)'de, d_e tahmin edilen eşitsizlik değerini temsil ederken, d_g ise doğru eşitsizlik değerini temsil etmektedir.

Yanlış eşleşen piksellerin yüzdesi stereo eşleştirme yöntemlerinde en çok kullanılan hata metriği olup Scharstein ve Szeliski (2002) tarafından önerilmiştir. Bu hata metriğine göre, elde edilen eşitsizlik haritasındaki değerler ile doğru eşitsizlik haritasındaki değerler

karşılaştırılmaktadır. Böylece referans bir eşik değeri seçilerek hatalı olan piksellerin yüzdeleri hesaplanmaktadır. Yanlış eşleşen piksellerin yüzdesi Denklem (2.3) yardımıyla hesaplanır.

$$B = \frac{1}{N} \sum_{i \in \Omega} (|d_e^i - d_g^i| > \delta_d) \times 100 \quad (2.3)$$

Denklem (2.3)'de, δ_d terimi belirlenen hata eşikini temsil etmektedir. Örneğin, 1 piksellik hata oranı hesaplanırken $\delta_d = 1$ olarak alınmaktadır.

Küresel fark ölçüm metodu farklı derinliklerde meydana gelen hataları incelemek amacıyla kullanılmaktadır. Bu ölçüm metodu özellikle otonom sistemler için üretilen eşitsizlik haritasının doğruluğunu değerlendirmede oldukça önemlidir. Derinlik bilgisinin kullanıldığı otonom sistemlerde yakındaki ve uzaktaki nesnelerin derinliklerinin doğrulukları ayrı ayrı önem taşımaktadır. Yakındaki nesnelerin derinlik bilgisinin yüksek doğrulukta olması çarpışma önleme sistemleri için hayati önem taşımaktadır. Uzaktaki nesnelerin derinlik bilgilerinin doğru olması ise rota belirleme işlemlerinde önem taşımaktadır.

Küresel fark doğru derinlik değeri ile tahmin edilen derinlik değeri arasındaki hatayı ölçmek için tasarlanan mesafeye duyarlı bir hata ölçüm metriğidir. Mesafe duyarlılığının hesaplanmasında kullanılan derinlik bilgileri; eşitsizlik değeri, stereo kameraların odak uzunluğu ve taban çizgisi uzunluğundan faydalanan trigonometrik yöntemlerle elde edilmektedir.

GD hata ölçütü hesaplanırken ilk adım olarak derinlik haritası, derinlik eksenini boyunca K noktaya bölünmektedir ve örneklenen her k derinlik noktası için R_k ölçüm aralığı, $[k - r, k + r]$ olarak tanımlanmaktadır. Tüm olası derinlik değerlerinde oluşabilecek eşitsizlik değerindeki sapmaların ayırt edilmesinde belirlenen k ölçüm aralığı referans alınmaktadır. Bu ölçüm aralığına göre mutlak bağıl fark (Absolute Relative Difference, ARD) değerleri hesaplanmaktadır. ARD değerleri Denklem (2.4)'de belirtildiği gibi elde edilmektedir.

$$ARD_k = \frac{1}{N_{R_k}} \sum_{y_g \in R_k} \frac{|y_e - y_g|}{y_g} \quad (2.4)$$

Denklem (2.4)'de y_e , y_g ve N_{R_k} terimleri sırasıyla tahmin edilen derinlik değerini, doğru derinlik değerini ve R_k içerisindeki toplam piksel sayısını temsil etmektedir.

GD metriği, her örnekteki (k) ARD değerlerinin ortalaması alınarak Denklem (2.5)'de gösterildiği gibi elde edilmektedir.

$$GD = \frac{1}{K} \sum_{k \in K} ARD_k \quad (2.5)$$

B ve EPE ölçüm metriklerinde hata değerleri 2 boyutlu (2B) düzlemdeki eşitsizlik haritaları kullanılarak elde edilirken, GD metriğinde ise hata değerleri 3 boyutlu (3B) düzlemdeki derinlik haritaları kullanılarak elde edilmektedir.

2.4 Nicel ve Nitel Karşılaştırmalar

Bu tez kapsamında farklı veri kümelerinde yapılan nicel karşılaştırmalarda elde edilen sonuçlar çizelgeler halinde sunulmaktadır. KITTI 2012 veri kümesindeki test sonuçlarının verildiği çizelgelerde eşitsizlik değeri bilinen bütün piksellerin hesaplamaya katıldığı durum “all” ile ifade edilirken, doğru eşitsizlik haritasında bulunan işaretlenmiş perdeleme (occlusions) içermeyen alanlardaki eşitsizlik değeri bilinen piksellerin hesaba katıldığı durum ise “noc” ile ifade edilmektedir. Ayrıca çizelgelerde hata değerleri 5, 4, 3 ve 2 gibi farklı piksel (px) değerleri için yüzdesel olarak verilmektedir.

KITTI 2015 veri kümesindeki test sonuçlarının verildiği çizelgelerde de “noc” ve “all” KITTI 2012 veri kümesine benzer anlama sahiptir. KITTI 2015 veri kümesindeki test sonuçlarının verildiği çizelgelerde D1 ifadesi ise 3 piksel hata oranını yüzdesel olarak göstermektedir. D1 ifadesinin yanında yer alan ifadeler ise görüntü içindeki sahne sınıflandırmalarını göstermektedir. Bu ifadelerden “fg” ön plan sahne alanını, “bg” arka plan sahne alanını ve “all” ise eşitsizlik değeri bilinen bütün pikselleri temsil etmektedir.

Scene Flow veri kümesinden elde edilen sonuçların verildiği çizelgelerde EPE hata

değerleri verilmektedir. Bu hata değerleri eşitsizlik değeri bilinen bütün piksellerin kullanılmasıyla hesaplanmaktadır.

Nitel karşılaştırmalar ise verilen eşitsizlik haritaları üzerinde yapılan görsel değerlendirmeler ile gerçekleştirilmiştir. Bu görsellerde hataların daha net bir şekilde gösterilmesini sağlamak amacıyla eşitsizlik haritalarının yanında hata haritaları da verilmektedir. Hata haritaları ait oldukları veri kümelerine göre farklı özellikler taşımaktadır.

Scene Flow veri kümelerinde verilen hata haritalarında renkler siyah-beyaz olarak verilmektedir. Siyah renkler belirlen eşik değere göre yanlış eşitsizlik değerini temsil ederken beyaz renkler ise doğru eşitsizlik değerine sahip pikselleri temsil etmektedir (Şekil 5.3, Şekil 6.4, Şekil 6.5).

KITTI 2012 veri kümesinde verilen hata haritalarında renkler gri seviyede verilmektedir. Siyah renkler belirlen eşik değere göre doğru eşitsizlik değerini temsil ederken beyaz renkler ise yanlış eşitsizlik değerine sahip pikselleri temsil etmektedir. Gri tonlardaki pikseller ise belirlenen eşik değere hata oranına göre siyah ve beyaza yakınlık tonlarıyla hata oranlarını göstermektedir (Şekil 4.7, Şekil 4.9, Şekil 6.6).

KITTI 2015 veri kümesindeki hata haritalarında ise piksel hataları görselleştirilirken logaritmik renk haritası kullanılarak hatanın büyüklüğüne göre farklı renklerde gösterimler türetilmiştir. Hata haritalarındaki renk kodlarının ölçeği oluşturulurken soğuk renkler (açık mavi) düşük hatayı sıcak renkler (koyu kırmızı) ise hatanın büyük olduğunu göstermektedir (Şekil 3.3, Şekil 4.8, Şekil 4.10, Şekil 6.7).

2.5 Deney Sistemi

Bu tez çalışması kapsamında PyTorch çatısı (framework) kullanılarak Python programlama dili ile dört farklı ağ modeli tasarlanmış ve başarımları incelenmiştir. Geliştirilen ağ modellerinin eğitim ve test süreçlerinde iki adet Nvidia GeForce RTX 2080 TI ekran kartı kullanılmıştır. Belirtilen çalışmalar Ubuntu 16.04 işletim sistemine, AMD Ryzen 7 1700 @3.0GHz işlemci ve 32 GB RAM belleğe sahip bir donanım üzerinde gerçekleştirilmiştir.

BÖLÜM III

EŞLEŞME MALİYETSİZ EŞİTSİZLİK HARİTASI DÜZELTME YÖNTEMİ

Eşitsizlik haritalarını düzeltmek amacıyla yapılan çalışmalar incelendiğinde eşleşme maliyetleri kullanılarak yapılan düzeltme işlemleri, eşleşme maliyetleri kullanılmadan yapılan düzeltme işlemlerine göre daha başarılı sonuçlar vermektedir (Cheng vd., 2018; Kim vd., 2019). Fakat bazı durumlarda eşleşme maliyetlerine doğrudan ulaşmak mümkün olmamaktadır. Eşleşme maliyetlerine ulaşmanın mümkün olmadığı durumlar, ZED (Stereolabs, San Francisco, ABD) ve Intel RealSense (Intel, Santa Clara, ABD) gibi bazı hazır sistemler de karşımıza çıkmaktadır. Bu sistemlerden elde edilen eşitsizlik haritaları hatasız değildir ve donanım, kullanıcıya sadece stereo görüntüleri ve bu görüntülere ait eşitsizlik haritalarını sunmamaktadır.

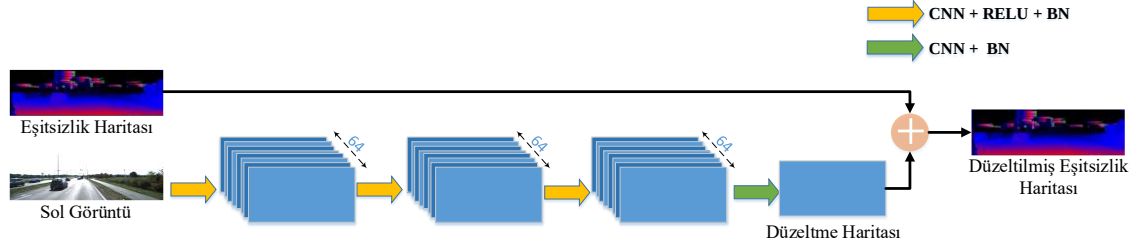
Bu bölümde eşleşme maliyetlerine ulaşmanın mümkün olmadığı durumlarda eşitsizlik haritası düzeltmek amacıyla sadece stereo görüntüler ve eşitsizlik haritaları ile çalışan derin öğrenme tabanlı bir ağ yapısı önerilmiştir ve bu ağ yapısı SimRef (Simple Refinement, Basit Düzeltme) olarak adlandırılmıştır. Önerilen SimRef ağı ile sadece düzeltme işlemi yapmaktadır. Bu işlem sırasında ise çalışma süresi mümkün olduğu kadar düşük tutulmaya çalışılmıştır.

Deneyisel çalışmalar KITTI 2012 ve KITTI 2015 veri kümeleri kullanılarak gerçekleştirilmiştir. Deneyisel çalışmaların sonunda önerilen SimRef ağının giriş olarak aldığı eşitsizlik haritasının doğruluğunu artırdığı gözlemlenmiştir. Elde edilen sonuçlar hem nitel hem de nicel karşılaştırmalar kullanılarak desteklenmiştir. Bu bölümde yapılan çalışmalar Ulakbilim kapsamında taranan bir dergide yayınlanmıştır (Emlek ve Peker, 2020).

3.1 Ağ Yapısı

Eşitsizlik haritalarının düzeltilmesi işleminde eşleşme maliyetleri kullanılarak yapılan çalışmalar daha başarılı sonuçlar vermektedir (Cheng vd., 2018; Kim vd., 2019). Fakat bu yöntemler doğrudan eşitsizlik haritaları veya stereo görüntüler üzerinden yapılan düzeltme işlemleri ile karşılaştırıldığı zaman işlem yükü bakımından oldukça ağır

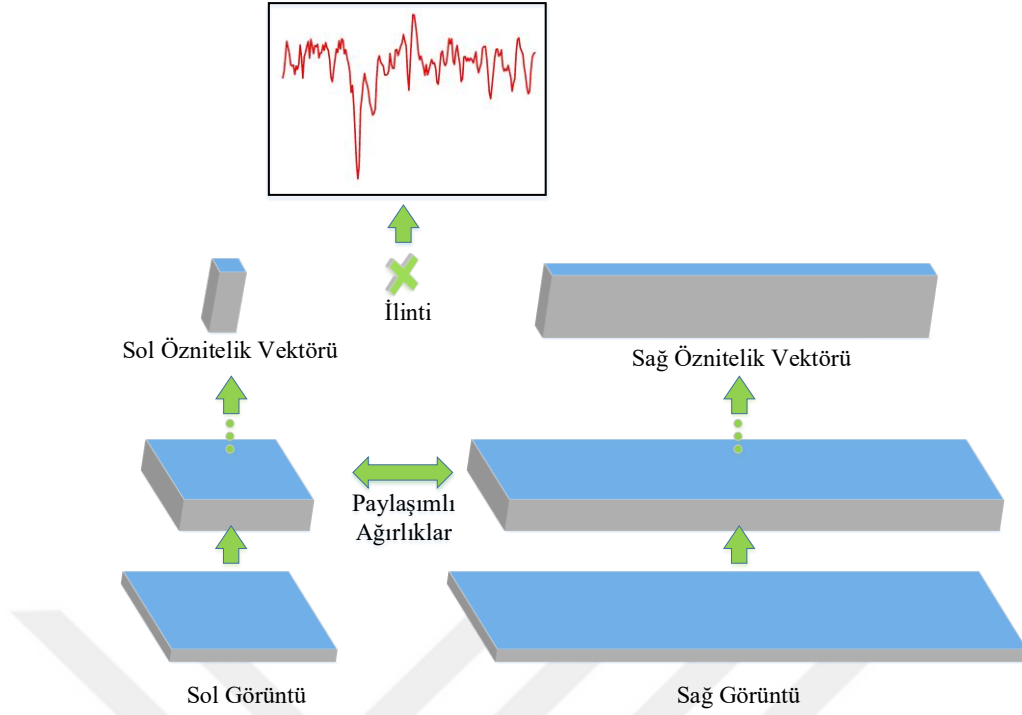
kalmaktadır. Hem işlem yükü hafifliği hem eşleşme maliyetlerine erişilebilirlikler göz önünde bulundurularak eşitsizlik haritalarının düzeltilmesi amacıyla bir ağ tasarlanmıştır. Tasarlanan SimRef ağ yapısı Şekil 3.1’de verilmektedir.



Şekil 3.1. Eşitsizlik haritalarının düzeltilmesi amacıyla önerilen SimRef ağ yapısı (Emlek ve Peker, 2020)

SimRef ağ yapısı giriş olarak stereo görüntülerden sol kameradan elde edilen görüntüyü ve eşitsizlik haritasını kullanmaktadır. Eşitsizlik haritası ise literatürde kabul görmüş olan Content-CNN (Luo vd., 2016) ağından elde edilen eşleşme maliyetlerinden elde edilmiştir.

Content-CNN ağ yapısı Şekil 3.2’de verilmiştir. Bu ağ yapısı stereo görüntüler kullanarak eşleşme maliyetleri elde etmektedir. Content-CNN ağ yapısı sol görüntü üzerinden aldığı 9×9 ’luk görüntü parçasına karşın sağ görüntüden muhtemel tüm arama düzlemini giriş olarak almaktadır. Bu giriş görüntüleri siyam ağlara verilerek paylaşılan ağırlık değerleri ile sıralı evrişim işlemleri uygulanmıştır. Evrişim işlemlerinde ise çıkış öz niteliklerinin boyutu her bir adımda hem genişlik hem yükseklik olarak iki azaltılmaktadır. Siyam ağların sonunda sol giriş görüntüsünden $1 \times 1 \times N$, sağ görüntüden ise $1 \times D \times N$ boyutunda öz nitelikler elde edilmektedir. Bu boyut ifadelerindeki D maksimum eşitsizlik değerini, N öz nitelik sayını temsil etmektedir. Elde edilen öz niteliklerden ise ilinti işlemi ile benzerlik puanları hesaplanmaktadır. Bu yapılan işlemler görüntü üzerindeki tüm pikseller için gerçekleştirildiğinde eşleşme maliyetleri elde edilmektedir. Luo vd. (2016) çalışmalarında Content-CNN ağı ile elde ettikleri eşleşme maliyetlerini kullanarak geleneksel stereo eşleştirme algoritmaları yardımıyla eşitsizlik haritaları elde etmişlerdir.



Şekil 3.2. Content-CNN ağ yapısı (Luo vd., 2016)

SimRef ağı ile önerilen düzeltme işleminde Content-CNN ağından elde edilen eşleşme maliyetlerine “kazanan-hepsini-alır (Winner-take-all - WTA)” yöntemi uygulanarak elde edilen eşitsizlik haritaları kullanılmaktadır. Kazanan-hepsini-alır yöntemi eşleşme maliyetlerinden en düşük değerin indisini seçme işlemidir ve Denklem (3.1) verilmektedir.

$$d_p = \underset{d}{\operatorname{argmin}} C_{Content_CNN}(p, d) \quad d \in [0 \ d_{max}] \quad (3.1)$$

Denklem (3.1)’de Content-CNN ağından üretilen eşleşme maliyetleri $C_{Content_CNN}$ ifadesi ile gösterilmektedir. p ifadesi sol görüntüdeki her bir pikseli, d ise eşitsizlik değerini temsil etmektedir. d eşitsizlik değerinin aralığı ise 0 ile maksimum eşitsizlik (d_{max}) olarak tanımlanmıştır.

Önerilen SimRef eşitsizlik haritası düzeltme ağı, giriş olarak aldığı sol kamera görüntüsüne sıralı evrişim işlemleri uygulamaktadır. Ağın eğitimini kolaylaştırmak ve başarımını artırmak amacıyla her bir evrişim katmanının sonunda yığın normalizasyonu (Batch Normalization, BN) işlemi uygulanmıştır. Ayrıca ağın doğrusallığını bozabilmek için son katmanda bulunan evrişim katmanı hariç diğer tüm evrişim katmanlarında ReLU (Doğrultulmuş Doğrusal Ünite, Rectified Linear Unit) aktivasyon fonksiyonu kullanılmıştır.

SimRef ağındaki evrişim işlemlerinde kullanılan çekirdek boyutları 5×5 olarak belirlenirken öznitelik sayısı ise 64 olarak seçilmiştir. Evrişim işlemlerinin son katmanında öznitelik sayısı 1 olarak seçilmiştir. Son katmandaki bu işlemin sonucunda düzeltme haritası elde edilmiştir. Düzeltme haritası ile giriş eşitsizlik haritasının birleştirilmesi ile nihai eşitsizlik haritası elde edilmiştir.

Ayrıca en uygun ağ yapısının belirlenmesi amacıyla farklı sayıda CNN'ler kullanılarak eğitim ve test işlemi gerçekleştirilmiştir. CNN katman sayısı için üç, dört ve beş olmak üzere üç farklı boyut belirlenmiştir. CNN katman sayısı boyutları seçilirken işlem süresinin fazla artmaması göz önünde tutularak düşük değerler seçilmeye çalışılmıştır. Böylece ağı ileride tasarlanacak gerçek zamanlı bir uçtan uca ağ yapısına veya gerçek zamanlı bir sisteme dâhil edilmesine olanak sağlanmıştır.

3.2 DeneySEL Sonuçlar

SimRef ağının eğitim ve test işlemleri KITTI 2012 (Geiger vd., 2012) ve KITTI 2015 (Menze ve Geiger, 2015) veri kümelerindeki eğitim kümesi kullanılarak gerçekleştirilmiştir. Eğitim işlemi için veri kümelerindeki görsellerden %80'i, doğrulama işlemleri için %10'u ve test işlemleri içinde geriye kalan %10'luk kısım seçilmiştir.

Ağın öğrenilebilir parametrelerinin başlangıç değerlerini belirlemek için Xavier yöntemi kullanılmıştır. Eğitim işlemi esnasında kayıp (yitim, loss) fonksiyonu olarak Denklem (3.2)'de verilen $L1$ regresyon fonksiyonu kullanılmıştır.

$$E_{kayıp} = \frac{1}{N} \sum_{n=1}^N (|Y_{gerçek} - Y_{tahmin}|) \quad (3.2)$$

Bu denklemdeki ifadelerden $Y_{gerçek}$ doğru eşitsizlik haritasını, Y_{tahmin} ise SimRef ağının sonunda üretilen eşitsizlik haritasını temsil etmektedir.

Denklem (3.2)'de verilen kayıp fonksiyonunu en aza indirmek için ADAM yöntemi kullanılmıştır. ADAM yöntemine ait parametreler sırayla; öğrenme oranı 0.001, beta değerleri ise 0.9 ve 0.999 olarak seçilmiştir. Ağın eğitimi ise tüm veri kümelerinde 100 tekrar (epoch) olarak gerçekleştirilmiştir.

SimRef ağı için en uygun yapının belirlenmesi amacıyla, farklı sayıda CNN kullanılarak üç farklı veri kümesinde eğitim ve test işlemleri gerçekleştirilmiştir. Farklı sayıdaki CNN'lerden farklı eğitim kümelerinde elde edilen 3 piksel hata oranları Çizelge 3.1'de verilmiştir. Çizelgede uygulanan CNN katman sayıları parantez içerisinde belirtilmektedir.

Çizelge 3.1. SimRef ağ yapısında farklı CNN katman sayıları kullanılarak elde edilen 3 piksel hata oranları (Emlek ve Peker, 2020)

Yöntem	Veri Kümelerinde Oluşan Hatalar (%)		
	KITTI 2012	KITTI 2015	KITTI 2012 + KITTI 2015
Content-CNN	8,93	6,23	7,73
SimRef (3)	8,88	5,94	7,57
SimRef (4)	8,83	5,89	7,54
SimRef (5)	8,83	5,87	7,52

Çizelge 3.1'de görüldüğü gibi ilk yöntem olarak, ağın girişine verilen başlangıç eşitsizlik haritasının kaynağı olan Content-CNN ağı verilmiştir. Daha sonra CNN katman sayısına göre sıralı bir şekilde SimRef ağ yapıları kullanılarak elde edilen sonuçlar verilmektedir. Bu ağların eğitim ve test kümeleri de sırayla KITTI 2012, KITTI 2015 ve KITTI 2012+KITTI 2015 olarak belirtilmiştir. KITTI 2012+KITTI 2015 veri kümesi ise her iki veri kümesinin birleştirilmesi ile elde edilmiştir.

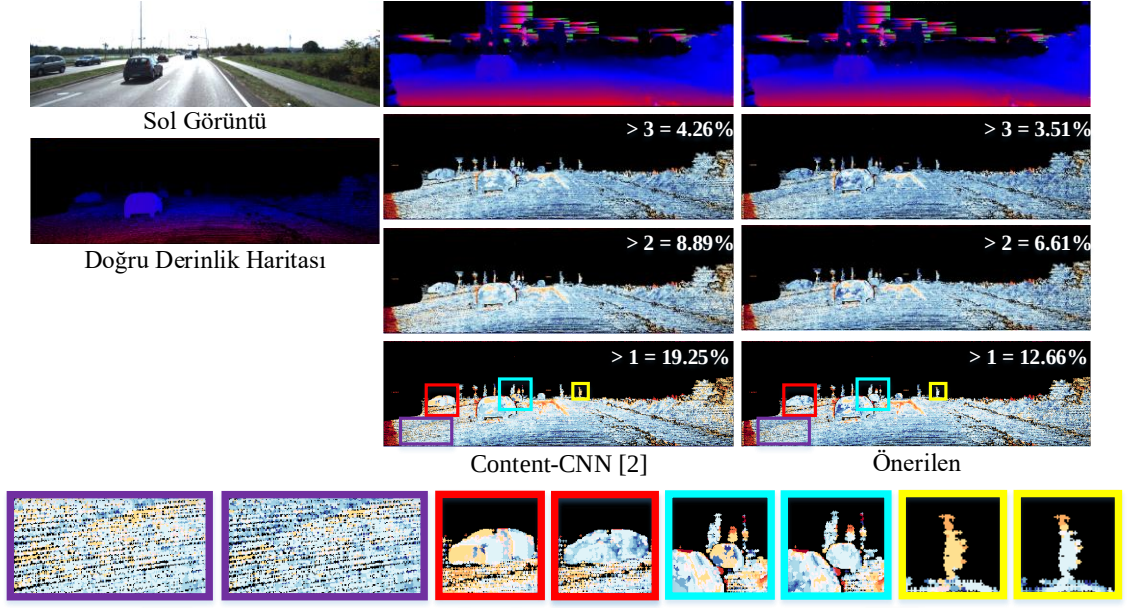
Referans olarak ağı verilen eşitsizlik haritaları ve elde edilen eşitsizlik haritaları, doğru eşitsizlik haritaları ile kötü piksel yüzdesi hata oranı kullanılarak karşılaştırılmıştır. Kötü piksel yüzdesi hata oranı eşik değeri 3 piksel olarak seçilmiştir. Çizelge 3.1'de görüldüğü gibi SimRef ağı, referans olarak alınan Content-CNN ağı ile karşılaştırıldığında bütün farklı CNN'li yapılarda başarıyı artırdığı gözlemlenmektedir. Aynı zamanda CNN katman sayısındaki artışında başarıyı artırdığı gözlemlenmiştir.

Üç piksel hata oranları ile verilen nicel karşılaştırmalara ilaveten veri kümeleri üzerindeki görüntüler kullanılarak nitel karşılaştırmalar da yapılmıştır. Bu karşılaştırmalardan elde edilen sonuçlar, iyileştirmelerin daha çok hangi alan ve koşullarda gerçekleştiğini göstermektedir.

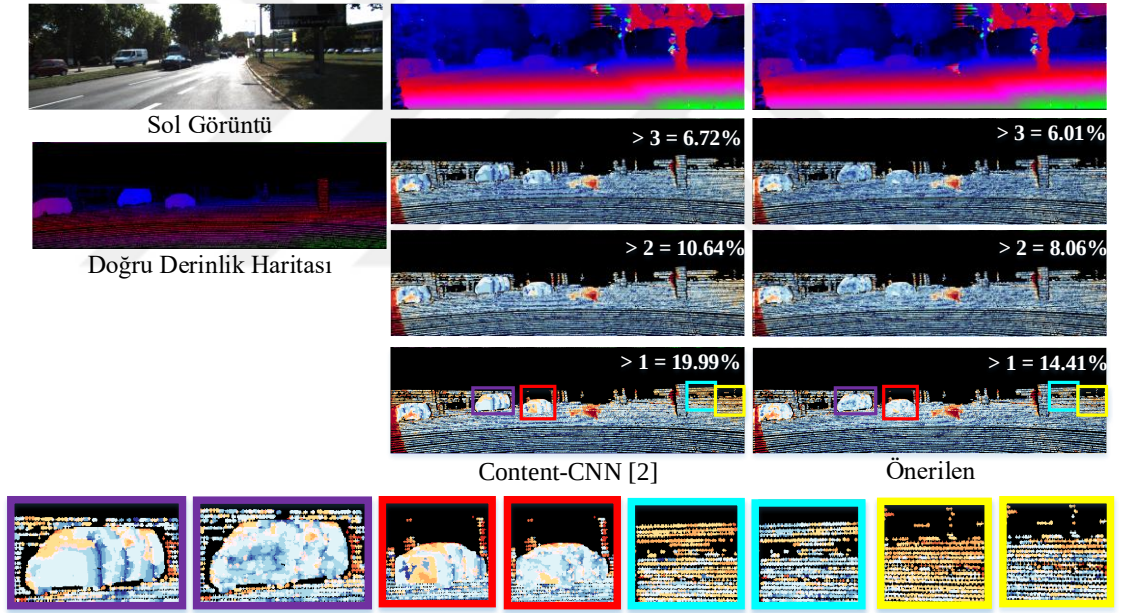
Content-CNN ağından elde edilen eşitsizlik haritaları ve SimRef (4) ağ yapısı kullanılarak elde edilmiş düzeltilmiş eşitsizlik haritalarının nitel karşılaştırmaları Şekil 3.3'te verilmiştir. İlk sütunda sol görüntü ve doğru eşitsizlik haritası verilirken, diğer sütunlarda ise sağa doğru sıra ile Content-CNN ve SimRef ağından elde edilen sonuçlar verilmiştir. Her bir eşitsizlik haritasına ait 3, 2 ve 1 piksellik hata haritaları alt alta verilmektedir. Şekil 3.3'te a ve b örneklerine ait karşılaştırmaya görüntülerinin en alt satırında ise hata haritalarının içerisinden seçilen alanların büyütülmüş görselleri yer almaktadır. Bu görsellerden soldaki görüntüler Content-CNN ağından elde edilen eşleşme maliyetlerine WTA uygulanarak elde edilen eşitsizlik haritalarına aittir. Sağ taraftaki görüntüler ise önerilen SimRef ağı kullanılarak elde edilmiştir. Büyütülerek verilen bu alanlar karşılaştırma işleminde dikkate değer farklılıkların olduğu bölgeleri göstermektedir.

Şekil 3.3'te görüldüğü üzere, giriş eşitsizlik haritası ile SimRef ağ yapısı kullanılarak elde edilen sonuçlarda 3-piksel ve 2-piksel hataları karşılaştırıldığı zaman, elde edilen eşitsizlik haritasında gözle görülür iyileştirmeler olduğu görülmektedir. En alt satırda verilen 1-piksel hatalarında ise iyileşmeler daha net bir şekilde gözlemlenmektedir. Özellikle 1-piksel hatalarında, herhangi bir desen içermeyen alanlar ve nesnelerin köşelerinde yer alan düzeltmeler dikkat çekmektedir. Şekil 3.3'te verilen görüntülerde araba, trafik işaretleri gibi nesnelerin hem kenar hem de yüzey alanlarındaki derinlik değerlerinin doğruluklarında dikkate değer bir artış olduğu gözlemlenmektedir. Bu doğruluk artışları eşitsizlik değerlerindeki süreklilik ve kenar bilgilerinin hassasiyetine ihtiyaç duyan üç boyutlu bölütleme ve üç boyutlu nesne tanımlama gibi uygulamalar için büyük bir avantaj sağlamaktadır.

Sonuç olarak, eşitsizlik haritalarının iyileştirilmesi amacıyla eşleşme maliyetleri kullanılmadan sadece sol kamera görüntüsü ve eşitsizlik haritasının giriş olarak kullanıldığı bir ağ yapısı önerilmiştir. Sol kamerada ki görüntülerden elde edilen öznitelikler eşitsizlik haritalarının iyileştirilmesinde kullanılmıştır. Nitel ve nicel



(a)



(b)

Şekil 3.3. KITTI 2015 veri kümesinde Content-CNN ve SimRef ağ yapısı kullanılarak elde edilen eşitsizlik haritalarının karşılaştırılması (Emlek ve Peker, 2020)

karşılaştırmalar sonucunda Content-CNN ağından elde edilen eşitsizlik haritalarının, çok ağır bir işlem yükü oluşturmadan gerçek zamanlı sistemlere entegre edilebilecek şekilde eşitsizlik haritalarının iyileştirildiği gösterilmiştir.

Önerilen SimRef ağ yapısında işlem yükünün az olmasının temel sebebi önerilen ağ yapısının oldukça sığ (shallow) olmasıdır. Ağ bu hafifliği sayesinde, literatürde herhangi

bir yöntemle eşitsizlik haritası elde edilebilen yöntemlere yardımcı olarak kullanılabilir. Ayrıca SimRef yöntemi, mevcut stereo görüntü tabanlı eşitsizlik bilgisini kullanan sistemler içerisinde performansı arttırmak için doğrudan kullanılabilir. Böylece, stereo görüntülerin kullanıldığı ve eşitsizlik bilgisine ihtiyaç duyulan uygulamalarda da başarımların artırılmasına olanak sağlanacaktır.



BÖLÜM IV

YİNELENEN SİNİR AĞI TABANLI EŞLEŞTİRME MALİYETİ İYİLEŞTİRME AĞI

Stereo eşleştirme alanında ki çalışmalar her ne kadar gelişim gösterse de mükemmel eşitsizlik haritasına ulaşmak günümüzde hala mümkün değildir. Eşitsizlik değerinin doğruluğunu etkileyen en önemli faktör ise eşleşme maliyetlerinin doğruluğudur. Eşleşme maliyetleri, yatay eksen boyunca komşu piksellerin piksel değerlerinin eğilimleri hakkında bilgi paylaşmaktadır.

Bu bölümde eşleşme maliyetlerinin paylaştığı bu sıralı bilgileri kullanarak eşitsizlik haritası elde edilirken oluşan yanlış eşitsizlik değerlerinin düzeltilmesini mümkün kılacak bir ağ yapısı önerilmiştir. Bu ağ yapısında piksellerin ve dolayısıyla eşleşme maliyetlerinin sıralı yapısı kullanılarak eşleşme maliyetlerinin düzeltilmesi amaçlanmıştır. Bu düzeltme işleminin gerçekleştirilebilmesi için literatürde daha önce kullanılmayan şekilde pikseller arasında sıralı bilgiler kullanılmıştır. Önerilen ağ yapısı, yatay eksen boyunca komşu piksellerin birbirleri ile paylaştığı piksel bilgilerini kullanmaktadır. Komşu pikseller arasında paylaşılan sıralı bilgilerin kullanılması için yinelenen sinir ağları (Recurrent Neural Network, RNN) kullanılmıştır. RNN yapısı kullanılarak eşleşme maliyetlerinin düzeltilmesi için önerilen ağ yapısına Yinelenen Sinir Ağı Tabanlı Maliyet İyileştirme Ağı (Cost Refinement Recurrent Neural Network, CR-RNN) adı verilmiştir.

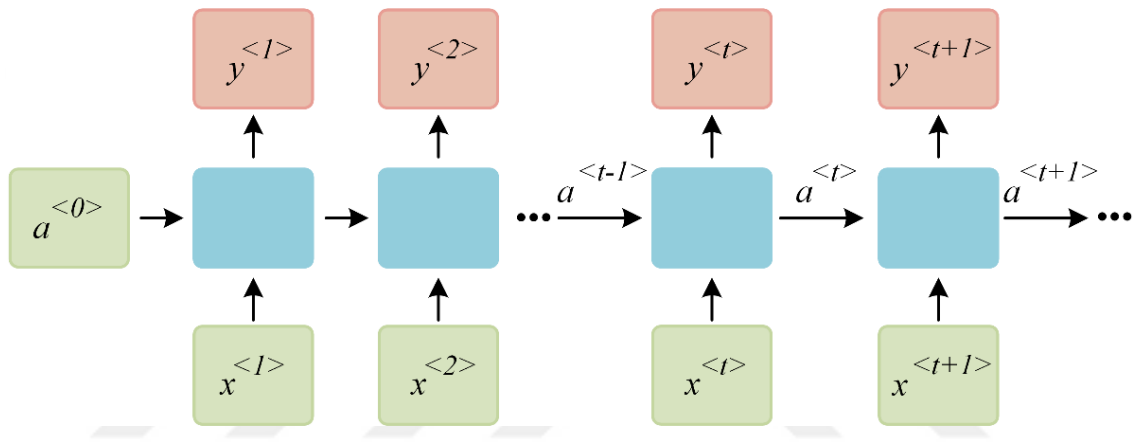
CR-RNN girişi olan eşleşme maliyetlerini elde etmek için hali hazırda kullanılan ve kabul görmüş yöntemler seçilmiştir. Üç farklı yöntemle elde edilen eşleşme maliyetleri CR-RNN yapısına uygulanmıştır. İki farklı veri kümesinden elde edilen sonuçlar, eşleşme maliyetlerinin sıralı yapısı kullanılarak maliyet hacimlerinin doğruluğunun artırılabilirliğini göstermiştir. Bu bölümde yapılan çalışmalar SCI kapsamındaki bir dergide yayınlanmıştır (Emlek ve Peker, 2021a).

4.1 Ağ Yapısı

CR-RNN yapısının amacı, eşitsizlik haritalarının elde edilmesinde en önemli paya sahip olan eşleşme maliyetlerinin doğruluğunu pikseller arasındaki sıralı bilgilerden

yararlanarak arttırmaktır. Pikseller arasındaki ilişkilerden yararlanarak eşleşme maliyetlerinin kalitesini artırmak amacıyla ses işleme problemlerinde olduğu gibi sıralı bilgilerin işlenmesi için yinelenen sinir ağıları kullanılmıştır.

RNN'ler, her zaman adımında giriş ve çıkış verileri arasındaki bağımlılığı depolamak ve kullanmak için bellek olarak gizli bir katman kullanmaktadır. RNN yapısını bu çalışmaya uyarlayabilmek amacıyla ardışık piksel değerleri, RNN'nin zaman adımları olarak kabul edilmiştir. Böylece, ardışık pikseller tarafından taşınan sıralı bilgilerin işlenmesi sağlanmıştır. Tipik bir RNN yapısı Şekil 4.1'de gösterilmektedir.



Şekil 4.1. Tipik bir RNN yapısı

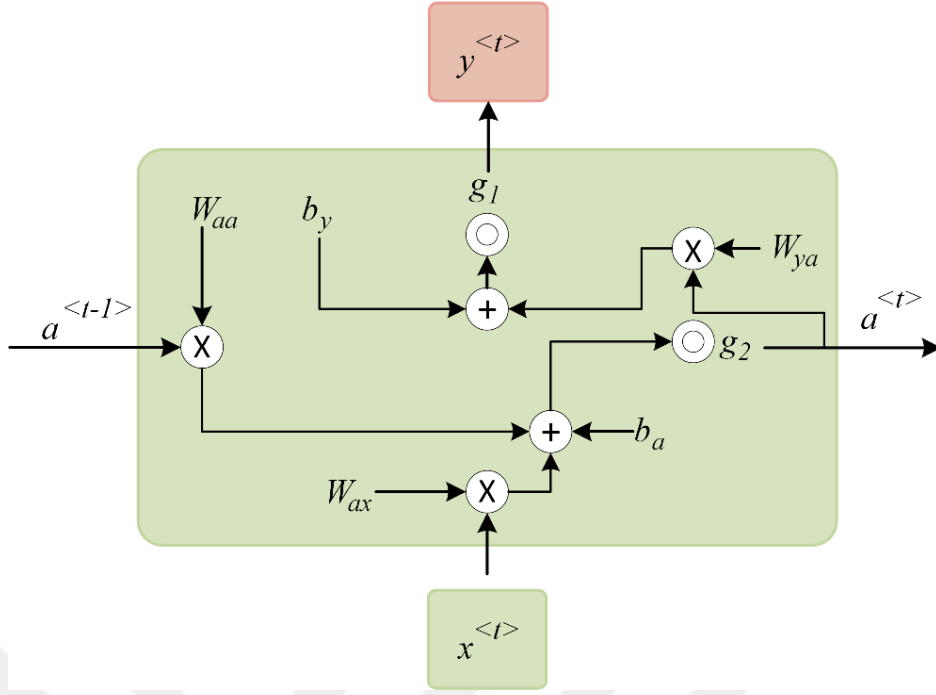
Şekil 4.1'de her bir t zaman adımı için $x^{<t>}$ giriş sinyali karşılık $y^{<t>}$ çıkış sinyali Denklem (4.1)'de belirtildiği gibi hesaplanmaktadır.

$$y^{<t>} = g_1(W_{ya}a^{<t>} + b_y) \quad (4.1)$$

Yukarıda verilen ifadede ki $a^{<t>}$ terimi Denklem (4.2)'de belirtildiği gibi hesaplanmaktadır.

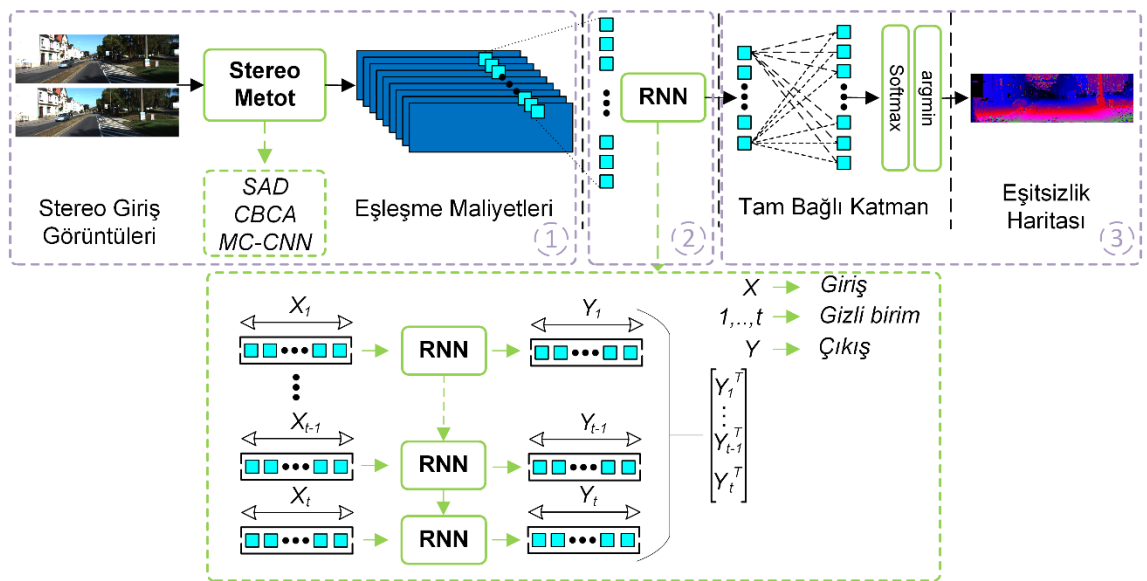
$$a^{<t>} = g_2(W_{aa}a^{<t-1>} + W_{ax}x^{<t>} + b_a) \quad (4.2)$$

Denklem (4.1) ve Denklem (4.2)'de belirtilen g_1 ve g_2 aktivasyon fonksiyonu, W_{aa} , W_{ax} , W_{ay} , b_a ve b_y terimleri ise geçici olarak paylaşılan ağırlıklardır. Denklem (4.1) ve Denklem (4.2)'de verilen ifadelerin görselleştirilmiş hali Şekil (4.2)'de verilmektedir.



Şekil 4.2. RNN hücresi

RNN yapıları, zaman adımı sayesinde sıralı bilgilerin geçmiş durumlarını inceleyerek bir sonraki durum hakkında tahmin yürütebilmektedirler. Önerilen ağda RNN yapısının eşleşme maliyetlerindeki sıralı bilgileri kullanarak eşleşme maliyetlerin doğruluğunu artırması amaçlanmıştır. Bu amaç doğrultusunda yinelenen sinir ağı tabanlı maliyet iyileştirme ağı (Cost Refinement Recurrent Neural Network - CR-RNN) yapısı önerilmiştir. Önerilen CR-RNN yapısı Şekil 4.3'te verilmiştir.



Şekil 4.3. Yinelenen sinir ağı tabanlı maliyet iyileştirme ağı (Cost Refinement Recurrent Neural Network - CR-RNN) yapısı (Emlek ve Peker, 2021a)

Şekil 4.3'te gösterildiği gibi, CR-RNN yapısı üç katmandan oluşmaktadır. İlk katmanda düzeltilmiş stereo görüntülerden eşleşme maliyet hacimleri elde edilmiştir. Stereo görüntüler, Şekil 4.3'te gösterildiği gibi, stereo metotların girdileri olarak kullanılmıştır. Stereo metotlardan eşleşme maliyetlerini elde etmek için literatürde sıklıkla kullanılan kabul görmüş yöntemler seçilmiştir. CR-RNN yapısının sadece belirli durumlarda değil genel olarak tüm eşleşme maliyetleri üzerinde çalışabildiğini göstermek amacıyla farklı yöntemlerle elde edilen eşleşme maliyetleri kullanılmıştır. Bu amaçla her biri farklı yapılara sahip üç farklı stereo eşleştirme yöntemi seçilmiştir. Stereo yöntemlerden elde edilen eşleşme maliyetlerinin boyutları $g \times y \times d$ olarak gösterilmiştir. $g \times y \times d$ ifadeleri sırasıyla girdi görüntülerinin genişliğini, yüksekliğini ve maksimum eşitsizlik değerini temsil etmektedir.

İkinci katmanda ise, bir pikselin arama düzlemindeki tüm eşitsizlik değerlerine karşı elde ettiği eşleşme maliyetlerini içeren maliyet vektöründen oluşan giriş, RNN ağına verilmiştir. Bir başka deyişle RNN ağının girişleri, stereo yöntemlerle elde edilen eşleşme maliyetlerinden bir piksel için seçilen vektördür. Fakat bu vektörler RNN ağına giriş olarak verilebilmesi için “ zaman_adımı $\times$ giriş_sayısı ” şeklinde yeniden boyutlandırılmıştır. RNN katmanının çıkış boyutları da giriş boyutları ile aynı olup “zaman_adımı $\times$ giriş_sayısı” olarak belirlenmiştir.

Eşleşme maliyetleri RNN ağlarına verilirken, yeniden boyutlandırma işleminde en uygun zaman adımı sayısının belirlenmesi amacıyla farklı değerlerde eğitim ve test işlemleri gerçekleştirilmiştir. Zaman adımı boyutu değiştirildiği durumda, RNN'lerin giriş boyutları maksimum eşitsizlik değerine göre ayarlanmıştır. RNN'lerin giriş boyutlarının maksimum eşitsizlik ve zaman adımı cinsinden matematiksel ifadesi ise Denklem (4.3)'te verilmiştir.

$$maksimum_eşitsizlik = zaman_adımı \times giriş_sayısı \quad (4.3)$$

Zaman adımı boyutları değiştirilerek, kullanılan RNN sayısı ve RNN'lerin giriş vektörlerinin boyutları belirlenmiştir. Stereo yöntemlerden elde edilen maliyetler (X), farklı zaman adımları $[1 \dots t-1, t]$ için $[X_1 \dots X_{t-1}, X_t]$ olacak şekilde yeniden boyutlandırılmıştır. Her girdi sıralı dönüş yapısına sahip RNN'lere verilmektedir ve her bir X_i girişi için Y_i çıktısı elde edilmektedir. Sıralı dönüş yapısı ise sıralı bir yapıya sahip

olan eşleşme maliyetleri arasındaki etkileşimi sağlamıştır. Her adımdan elde edilen çıktılar $[Y_1 \dots Y_{t-1}, Y_t]$ birleştirilerek tek bir vektöre (Y) dönüştürülmüş ve bir sonraki katmana verilmiştir.

Son katmanda ise RNN ağının çıkışı tam bağlı bir yapay sinir ağı katmanına verilerek her bir d eşitsizlik değeri için eşleme maliyetleri elde edilmiştir. Tam bağlı katmanın nöron boyutu, maksimum eşitsizlik değerine eşittir. Elde edilen eşleşme maliyetlerinden eşitsizlik haritası üretebilmek amacıyla ağın çıkışı Softmax fonksiyonuna verilmiştir. Son adımda, elde edilen eşleşme maliyetlerinden en küçük değerlerin indeksleri alınarak eşitsizlik değerleri seçilmiştir. Sonuç olarak, CR-RNN mimarisi, uygulanabilirlik açısından oldukça basittir.

4.2 Deneysel Sonuçlar

Bu bölümde ilk olarak CR-RNN ağına giriş olarak verilen stereo yöntemler incelenmiştir. İkinci olarak, ağların eğitiminde kullanılan veri kümeleri ve eğitim işlemine ait detaylar verilmiştir. Üçüncü başlık altında da ağın en başarılı ve en hızlı sonuçları verdiği parametreler belirlenmiştir. Son adımlarda ise nicel ve nitel karşılaştırmalar verilmiştir.

4.2.1 Stereo metotlar

Bu alt bölümde, CR-RNN'nin girişinde kullanılan mevcut stereo yöntemlerin ayrıntıları verilmiştir. Eşleşme maliyet hacmini elde etmek amacıyla, üç farklı mevcut popüler stereo algoritması kullanılmıştır. Bu yöntemler mutlak farkların toplamı (Sum Of Absolute Differences - SAD) (Scharstein ve Szeliski, 2002), çapraz tabanlı maliyet toplama (Cross-Based Cost Aggregation - CBCA) (K. Zhang vd., 2009) ve MC-CNN-Hızlı (Žbontar ve Lecun, 2015)'dir.

Mutlak farkların toplamı (SAD), görüntüler üzerinde önceden belirlenmiş dikdörtgen veya kare gibi belirli bir alandaki pikselleri karşılaştırarak eşleşme maliyetlerini hesaplayan basit bir yöntemdir. Belirlenen alan içindeki eşleşme maliyetleri hesaplanırken Denklem (4.4)'de tanımlanan ifade kullanılmaktadır.

$$C_{SAD} = \sum_{q \in N_p} |I_L(p) - I_R(q - d)| \quad (4.4)$$

Denklem (4.4)'de verilen ifadeye göre I_L ve I_R sırasıyla sol ve sağ stereo görüntü çiftlerini, N_p ise önceden tanımlanmış pencere içerisindeki q piksellerini içeren alanı temsil etmektedir. Bunlara ilaveten, d terimi eşitsizlik değerini belirtmektedir.

Çapraz tabanlı maliyet toplama (CBCA) yönteminde, her pikselin yalnızca bulunduğu nesnenin sınırları içinde uyarlanabilir bir pencere alanı belirlenmesi sağlanmaktadır. Bu uyarlanabilir pencere belirlenirken ise her bir pikselin konumunun çevresinde benzer görüntü yoğunluk değerlerine sahip piksellerden oluşan bir yerel alan seçilmektedir. Uyarlanabilir pencerelerin sınırları birbirini sıralı olarak takip eden iki adımda belirlenmektedir. İlk adımda, her piksel için dikey ekseninde piksel komşuluklarındaki yoğunluk değişimlerine göre yerel destek bölgesinin dikey sınırları tanımlanmaktadır. İkinci adımda ise belirlenen dikey eksenindeki her bir piksel için yatay eksenindeki komşu pikseller arasındaki yoğunluk farklılıkları dikkate alınarak uyarlanabilir pencerenin sınırları belirlenmektedir.

Bu çalışmada kullanılan üçüncü eşleşme maliyeti elde etme yöntemi olarak MC-CNN-Hızlı'dır. MC-CNN-Hızlı, aynı çalışmada sunulan doğru mimariye kıyasla daha düşük parametre ve daha hızlı çalışma süresine sahip olması sebebiyle seçilmiştir. MC-CNN-Hızlı ağı, stereo görüntülerden aldığı görüntü yamalarını karşılaştırarak her piksel için arama düzlemi boyunca eşleşme maliyetleri üretmektedir. Eşleşme maliyetleri Siyam ağ yapısı kullanılarak elde edilmektedir. Siyam ağların bir girişine sol görüntüden alınan görüntü yaması verilmektedir. Diğer girişine ise, arama düzlemi boyunca her bir eşitsizlik değeri için sağ görüntüden alınan farklı giriş görüntü yamaları verilmektedir. Eşleşme maliyetleri, Siyam ağının çıkışlarından elde edilen sol ve sağ öznitelik vektörlerinin noktasal çarpımları ile elde edilmektedir.

4.2.2 Veri kümeleri ve eğitim detayları

Önermiş olduğumuz CR-RNN modelinin eğitim ve test işlemleri hem KITTI 2012 (Geiger vd., 2012) hem de KITTI 2015 (Menze ve Geiger, 2015) veri kümeleri kullanılarak gerçekleştirilmiştir.

Eğitim aşamasında parametrelerin başlangıç değerlerinin belirlenmesinde Xavier yöntemi kullanılmıştır. Kayıp fonksiyonu olarak ağırlıklı çapraz entropi (Weighted Cross-Entropy, WCE) kaybı kullanılmıştır. Toplam kayıp değerini en aza indirmek için ise Nadam seçilmiştir. Başlangıç öğrenme oranı 0.001 ve beta değerleri sırasıyla 0.99 ve 0.999 olarak belirlenmiştir. Her yineleme için toplu iş boyutu 128'dir. Ağın eğitimi iki adet NVIDIA GTX 2080TI ekran kartı üzerinde 15 tekrar (epoch) gerçekleştirilmiştir.

Literatürde ağırlıklı çapraz entropi kaybı fonksiyonun kullanıldığı çalışmalar mevcuttur (Luo vd., 2016, Shaked and Wolf, 2017). Ağırlıklı çapraz entropi kaybı fonksiyonu Denklem (4.5)'deki gibi tanımlanmaktadır:

$$loss(y_i, y^{GT}) = \sum_{y_i} P(y_i, y^{GT}) \cdot \log \frac{e^{y_i}}{\sum_j e^{y_j}} \quad (4.5)$$

Denklem (4.5)'de y^{GT} doğru çıktı değerini belirtmektedir. $P(y_i, y^{GT})$ ise Denklem (4.6)'da verilen doğru değer etrafındaki hedeflenen yumuşak hedef dağıtımını (smooth target distribution) temsil etmektedir.

$$P(y_i, y^{GT}) = \begin{cases} \lambda_1 & \text{eğer } |y_i - y^{GT}| \leq 1 \\ \lambda_2 & \text{eğer } 1 < |y_i - y^{GT}| \leq 2 \\ \lambda_3 & \text{eğer } 2 < |y_i - y^{GT}| \leq 3 \\ 0 & \text{diğer durumlarda} \end{cases} \quad (4.6)$$

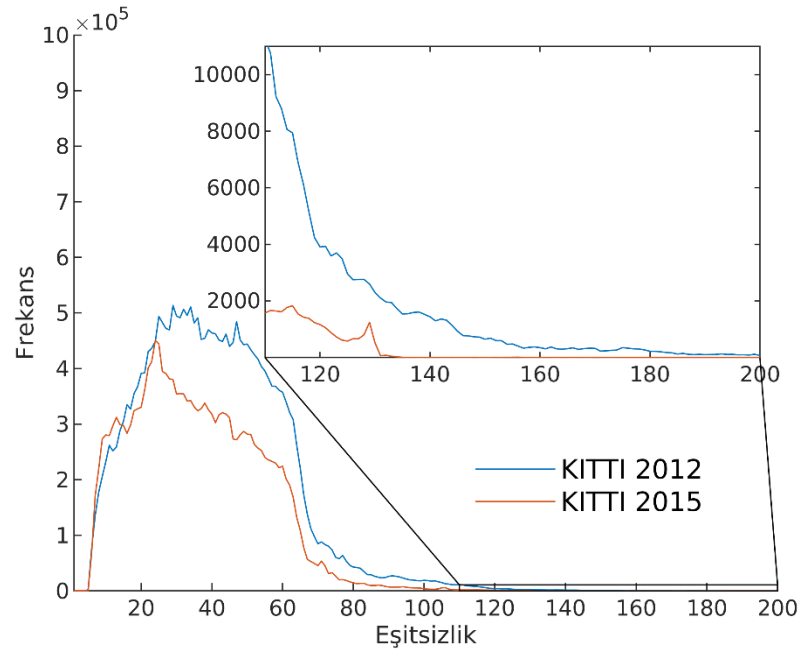
Yumuşak hedef dağıtımının parametreleri olan λ_1, λ_2 ve λ_3 terimlerinin değerleri sırasıyla 0.5, 0.2 ve 0.05 olarak ayarlanmıştır.

GD değerinin hesaplanması ve ARD eğrilerinin çizilmesinde ise farklı r değerleri kullanılmıştır. r değeri, GD hesaplamaları ve ARD eğrilerinin çizilmesi için sırasıyla 4 ve 0.5 olarak ayarlanmıştır. ARD eğrilerinde daha düzgün görselleştirmenin elde edilmesi amacıyla r değeri daha düşük olacak şekilde ayarlanmıştır.

CR-RNN yapısının eğitim kabiliyetini artırmak amacıyla veri artırma işlemi uygulanmıştır. Veri artırma (data augmentation) işlemi ağa verilen görüntülere bazı görüntü dönüşümleri uygulayarak görüntü çeşitliliğini ve sayısını artırmaktadır. Bu işlemin derin ağlarda daha iyi öğrenmeyi sağlayan önemli bir teknik olduğu

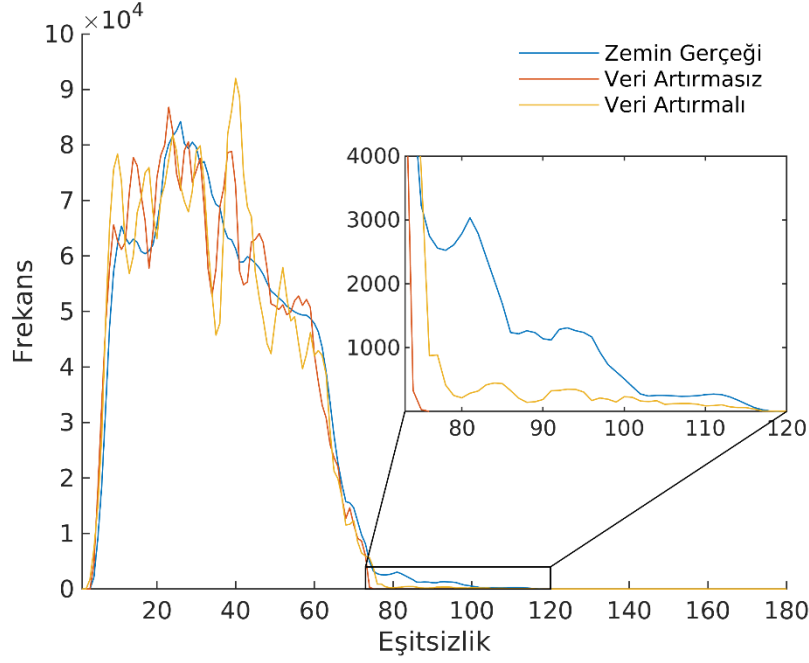
kanıtlanmıştır (Žbontar ve Lecun, 2015; Mayer vd., 2016; Yang vd., 2019).

KITTI 2012 ve KITTI 2015 veri kümelerindeki tüm eşitsizlik haritaları incelendiğinde, eşitsizlik değerlerinin büyük çoğunluğunun Şekil 4.4’de gösterildiği gibi 0 ile 100 değerleri arasında değiştiği görülmektedir. Bu durum, eğitim sırasında, ağırlık büyük oranda 0 ile 100 aralığındaki eşitsizlik bilgilerini öğrenmesine ve test işlemi sırasında genel olarak bu aralıktaki eşitsizlik değerlerini üretmesine sebep olmaktadır.



Şekil 4.4 KITTI 2012 ve KITTI 2015 veri kümelerinde yer alan eğitim setlerindeki eşitsizlik değerlerinin dağılımı (Emlek ve Peker, 2021a)

Eşitsizlik haritalarında gözlemlenen bu durumu ortadan kaldırmak amacıyla, eğitim sırasında kullanılan veriler $\text{max_eşitsizlik}/2$ ile max_eşitsizlik aralığında rastgele kaydırılarak genişletilmiştir. Veri artırma işlemi yapılarak ve veri artırma işlemi yapılmadan gerçekleştirilen eğitimlerin sonunda KITTI 2015 test veri kümesindeki eşitsizlik dağılımları Şekil 4.5’de verilmiştir. Şekil 4.5’de sunulan verilerin daha iyi anlaşılabilmesi için çekirdek boyutu 5 olan hareketli ortalama filtresi kullanılmıştır.



Şekil 4.5. KITTI 2015 test seti üzerinde veri artırma işlemi yapılarak ve yapılmadan elde edilen sonuçlarda eşitsizlik dağılımı (Emlek ve Peker, 2021a)

CR-RNN yapısının öğrenme yeteneklerini incelemek için, KITTI 2012 ve KITTI 2015 veri kümeleri iki farklı şekilde eğitim, doğrulama ve test veri kümesine bölünmüştür. Bölme işlemlerinde veri kümelerinde bulunan görüntüler belirlenen sayılara göre rastgele seçilmiştir. Bu aşamadaki test işlemlerinde sadece MC-CNN-Hızlı ağ yapısından üretilmiş eşleşme maliyetleri kullanılmıştır. Test işlemlerinin sonunda elde edilen sonuçlar Çizelge 4.1 ve Çizelge 4.2’de verilmiştir. Çizelge 4.1 ve Çizelge 4.2’deki sonuçların karşılaştırılmasında ise üç piksel hata oranları kullanılmıştır.

Çizelge 4.1. KITTI 2012 veri kümesinin farklı bölme oranlarında üç piksel test hata oranları (Emlek ve Peker, 2021a)

Eğitim-Doğrulama-Test (%)	B (%)>3px	
	MC-CNN-Hızlı	MC-CNN- Hızlı + CR-RNN
70-20-10	19.45	18.79
80-10-10	17.90	16.77

Çizelge 4.2. KITTI 2015 veri kümesinin farklı bölme oranlarında üç piksel test hata oranları (Emlek ve Peker, 2021a)

Eğitim-Doğrulama.-Test (%)	B (%)>3px	
	MC-CNN- Hızlı	MC-CNN- Hızlı + CR-RNN
70-20-10	15.54	14.70
80-10-10	13.77	12.15

Çizelge 4.1 ve Çizelge 4.2de’görüldüğü üzere eğitim setlerinin artan boyutu ile her iki veri kümesinde meydana gelen hata oranlarındaki azalış, eğitim veri kümelerinin daha büyük seçildiği durumlarda gerçekleşmiştir. Bu durum, daha iyi sonuçlar elde edilebilmesi için eğitim veri kümesinin daha büyük olması gerektiğini göstermektedir.

Bu bölümün devamında (Bölüm 4.2.3, 4.2.4 ve 4.2.5), eğitim, doğrulama ve test veri kümeleri sırasıyla %80, %10 ve %10 bölme oranı ile elde edilmiştir. Ayrıca elde edilen sonuçlar değerlendirilirken KITTI 2012 ve KITTI 2015 veri kümelerinin web sitesinde bulunan çevrimiçi test veri kümeleride kullanılmıştır. Verilen sonuçlarda çevrimiçi test veri kümelerinin kullanıldığı belirtilmediği takdirde, eğitim veri kümesi kullanılarak bölme işlemi ile elde edilen test veri kümesi kullanılmıştır.

4.2.3 Parametre eniyileme

Bu çalışmada kullanılan tekrarlayan sinir ağlarının yapısında ayarlanabilir iki ana parametre bulunmaktadır. Bu parametrelerden ilki, RNN'lerin gizli katman sayısıdır. Gizli katman sayısı doğrudan modelin parametre sayısını ve çalışma süresini değiştirmektedir. İkinci parametre ise RNN'nin zaman adımı boyutudur. Bu parametre, her iki veri kümesi için de 192 olan maksimum eşitsizlik değerine bağlıdır.

CR-RNN yapısının en uygun parametre değerlerini belirlemek amacıyla belirli değerler esas alınarak eniyileme yapılmıştır. Eniyileme işleminde bir parametreyi sabit bırakıp diğer parametrenin olası değerleri için elde edilen sonuçlara göre seçim yapılmıştır. İlk olarak, ağırlık, zaman adımı boyutu sabit tutulurken gizli katman sayısının değiştirilmesiyle eğitim ve test işlemleri gerçekleştirilmiştir. Uygun gizli katman sayısı belirlendikten sonra, zaman adımı boyutunun en uygun değerinin elde edilebilmesi için gizli katman sayısı sabit tutularak, farklı parametrelerde eğitim ve test işlemleri gerçekleştirilmiştir.

Ayrıca, bu bölümde kullanılan eşleşme maliyetleri MC-CNN-Hızlı ağ yapısından elde edilmiştir.

Eniyileme işleminin ilk aşamasında zaman adımı değeri iki kabul edilerek gizli katman sayısı değiştirilmiştir. Bu aşamada elde edilen sonuçlar Çizelge 4.3'de verilmektedir.

Çizelge 4.3. Zaman adımı değeri sabit tutularak, farklı gizli katman değerleri için elde edilen ağ performans değerleri (Emlek ve Peker, 2021a)

	Gizli Katman Sayısı				
	32	64	128	256	512
Parametre Sayısı	16,608	35,072	78,144	188,864	508,608
Hata Oranı B (%)>3px	11.97	11.35	11.28	11.23	11.30

Çizelge 4.3'e göre, gizli katman sayısının 256'dan 512'ye yükseldiği durum hariç, gizli katman sayısındaki artış yüzdesel B hata oranlarını azaltmaktadır. Buna ek olarak, gizli katman sayısının 256 dan 512'ye artışı parametre sayısında da önemli bir artışa sebep olmuştur. Gizli katman ve parametre sayısında ki bu artışın başarımlar üzerinde olumlu bir etkisi olmadığı gözlemlenmiştir. Bu nedenle gizli katman sayısı 256 olarak belirlenmiştir.

Gizli katman boyutu 256 olarak belirlendikten sonra zaman adım boyutunun belirlenmesi amacıyla eğitim ve test işlemleri yapılmıştır. KITTI 2012 ve KITTI 2015 veri kümelerinde farklı zaman adımı değerlerindeki 3 farklı hata metriğinde elde edilen değerler sırasıyla Çizelge 4.4 ve Çizelge 4.5 'de verilmiştir.

Çizelge 4.4. CR-RNN'nin farklı zaman adımı değerlerinde KITTI 2012 test seti üzerindeki başarımları (Emlek ve Peker, 2021a)

Zaman Adımı	Giriş Boyutu	Gizli Katman Boyutu	Parametre Sayısı	B (%)				EPE	GD
				>5px	>4px	>3px	>2px		
1	192	256	164,288	13.53	14.36	15.65	18.66	4.8488	26.84
2	96	256	188,864	13.29	14.19	15.52	18.47	4.9617	32.44
3	64	256	229,824	13.33	14.13	15.41	18.52	4.8131	31.04
4	48	256	274,880	13.21	14.11	15.37	18.19	5.0159	26.23
8	24	256	465,344	12.95	13.86	15.14	18.01	4.4572	26.17
12	16	256	659,904	12.89	13.78	15.13	18.26	4.4290	30.10

Çizelge 4.4'de görülebileceği gibi, KITTI 2012 veri kümesinde daha düşük karmaşıklığa sahip bir ağ oluşturmak için zaman adımı boyutunun düşük seviyelerde tutulması gerekmektedir. Aksi durumda, ağın karmaşıklığı önemli ölçüde artmaktadır. Çizelge 4.5'de verilen KITTI 2015 veri kümesindeki kötü eşleşen piksel hataları göz önüne alındığında, Çizelge 4.4'deki sonuçlara benzer durumlar gözlemlenmektedir. Ayrıca kötü eşleşen piksel yüzdelere göre daha iyi performanslı bir ağ elde etmek için zaman adımı değeri sayısının yüksek seçilmesi gerekmektedir. Sonuç olarak, ağ karmaşıklığı ve performansına ilişkin en iyi değer, soruna bağlı olarak belirlenmelidir.

Çizelge 4.5. CR-RNN'nin farklı zaman adımı değerlerinde KITTI 2015 test seti üzerindeki başarımı (Emlek ve Peker, 2021a)

Zaman Adımı	Giriş Boyutu	Gizli Katman Boyutu	Parametre Sayısı	B (%)				EPE	GD
				>5px	>4px	>3px	>2px		
1	192	256	164,288	9.24	9.95	11.24	14.17	3.1388	27.60
2	96	256	188,864	9.10	9.88	11.23	13.83	3.2980	21.26
3	64	256	229,824	9.31	10.04	11.24	13.85	4.5770	25.62
4	48	256	274,880	9.16	9.90	11.15	13.93	3.9793	24.13
8	24	256	465,344	9.04	9.80	11.13	13.92	4.0860	25.37
12	16	256	659,904	9.02	9.76	10.95	13.37	4.6316	22.23

Elde edilen eşitsizlik haritaları için nicel ve nitel karşılaştırmalar sırasıyla Bölüm 4.2.4 ve 4.2.5 de verilmektedir. Yapılan karşılaştırmalarda, hem performans (CR-RNN_P) hem de parametre boyutu (CR-RNN_C) açısından en iyi ağ yapıları kullanılmıştır. Niteliksel karşılaştırmalarda, daha iyi görselleştirme sağlanması amacıyla performans bazında en iyi ağ yapısı kullanılmıştır. Kullanılan bu ağ yapılarında, Çizelge 4.4 ve Çizelge 4.5 'den yararlanarak CR-RNN_P ağı için zaman adımı boyutu 12, CR-RNN_C ağına ise zaman adımı boyutu 1 olarak belirlenmiştir.

4.2.4 Nicel karşılaştırmalar

CR-RNN yapısının, eşleşme maliyetlerini iyileştirme performansını gösterebilmesi amacıyla üç farklı stereo metot kullanılmıştır. Bu stereo metotlar, SAD, CBCA ve MC-CNN-Hızlı olarak seçilmiştir. Karşılaştırma işlemlerinde ise bu yöntemlere ek olarak daha performanslı ve daha karmaşık bir ağ olan MC-CNN-Doğru yöntemi de kullanılmıştır.

Çizelge 4.6’da MC-CNN ağının ve önerilen CR-RNN ağı ile birleştirilen yapıların parametre sayılarının karşılaştırılması verilmiştir. SAD ve CBCA stereo metotları ise, ağ tabanlı çözümler olmadıkları için Çizelge 4.6’ya dahil edilmemişlerdir.

Çizelge 4.6. Ağ tabanlı yöntemlerin parametre boyutlarının karşılaştırılması (Emlek ve Peker, 2021a)

Mimariler	Toplam Parametre Sayısı
MC-CNN-Hızlı	111,424
MC-CNN-Doğru	870,449
CR-RNN _C	164.288
CR-RNN _P	659,904
MC-CNN-Hızlı + CR-RNN _C	275,712
MC-CNN-Hızlı + CR-RNN _P	771,328

Çizelge 4.6’ya göre, mimarilerin toplam parametre sayıları karşılaştırıldığında, önerilen her iki CR-RNN mimarisi de MC-CNN-Hızlı yapısından daha fazla parametre içerirken ve MC-CNN-Doğru yapısından daha az parametre içermektedir.

MC-CNN-Hızlı yapısı ile kullanılan CR-RNN ağları da MC-CNN-Doğru ağından daha az parametre içermektedir. Özellikle MC-CNN-Hızlı + CR-RNN_C yapısı MC-CNN-Doğru yönteminden oldukça düşük parametre sayısına sahiptir.

Farklı stereo metotlar kullanılarak CR-RNN ağlarının başarısı KITTI 2012 ve KITTI 2015 veri kümelerinde ki test setlerinde değerlendirilmiştir. Elde edilen sonuçların karşılaştırılmasında yüzdesel piksel hata oranları (%B) , EPE ve GD hata metrikleri kullanılmıştır. Bu hata oranları Çizelge 4.7 ve Çizelge 4.8’de verilmiştir.

Çizelge 4.7. CR-RNN ağ yapısının KITTI 2012 test setinde farklı stereo metotlarla karşılaştırılması (Emlek ve Peker, 2021a)

Mimari	B (%)				EPE	GD
	>5px	>4px	>3px	>2px		
MC-CNN-Hızlı	16.03	16.80	17.91	20.14	7.4006	35.12
MC-CNN-Doğru	13.18	13.81	14.76	16.86	6.5936	35.13
MC-CNN-Hızlı + CR-RNN _C	13.53	14.36	15.65	18.66	4.8488	26.84
MC-CNN-Hızlı + CR-RNN _P	12.89	13.78	15.13	18.26	4.4290	30.10
SAD	31.90	33.44	35.34	37.59	13.1638	74.15
SAD + CR-RNN _P	30.51	31.73	33.40	35.63	12.9390	72.47
CBCA	29.94	31.42	33.24	35.46	12.3728	68.13
CBCA+ CR-RNN _P	25.42	26.75	28.57	31.10	10.4674	50.10

Çizelge 4.8. CR-RNN ağ yapısının KITTI 2015 test setinde farklı stereo metotlarla karşılaştırılması (Emlek ve Peker, 2021a)

Mimari	B (%)				EPE	GD
	>5px	>4px	>3px	>2px		
MC-CNN-Hızlı	11.79	12.51	13.74	16.34	5.8471	35.46
MC-CNN-Doğru	9.65	10.21	11.21	13.67	5.1775	35.69
MC-CNN-Hızlı + CR-RNN _C	9.24	9.95	11.24	14.17	3.1388	27.60
MC-CNN-Hızlı + CR-RNN _P	9.02	9.76	10.95	13.37	4.6316	22.23
SAD	27.81	29.92	32.73	37.10	9.8349	62.01
SAD + CR-RNN _P	26.72	28.42	30.77	34.71	9.6686	60.74
CBCA	25.88	27.91	30.64	35.05	9.0861	56.12
CBCA+ CR-RNN _P	21.24	23.03	25.74	30.53	7.3726	45.43

Çizelge 4.7 ve Çizelge 4.8'de önerilen CR-RNN yapısının, giriş olarak aldığı bütün eşleşme maliyetleri için kullanılan üç hata metriğinde de doğruluğu artırdığı gözlemlenmektedir. Ağ tabanlı MC-CNN ile yapılan karşılaştırmalarda, MC-CNN-Hızlı + CR-RNN_C yapısı MC-CNN-Doğru'ya yakın performans gösterirken, MC-CNN-Hızlı + CR-RNN_P yapısı MC-CNN-Doğru'dan daha iyi performans göstermektedir. EPE ve GD hata metrikleri dikkate alındığında ise, MC-CNN-Hızlı + CR-RNN_C ve MC-CNN-Hızlı + CR-RNN_P, her iki MC-CNN (MC-CNN-Hızlı ve MC-CNN-Doğru) ağına kıyasla çok daha üstün performans göstermektedir.

Ayrıca, SAD + CR-RNN_P ve CBCA + CR-RNN_P'nin, MC-CNN örneğinde olduğu gibi orijinal eşleşme maliyetlerinden daha iyi performans sergilediği gösterilmiştir. Bu karşılaştırmalar, CR-RNN ağ yapısının hem 2B düzlemdeki eşitsizlik haritalarında hem

de 3B düzlemdeki derinlik haritalarında daha iyi sonuçlar verdiğini göstermektedir.

KITTI 2012 ve KITTI 2015 veri kümelerindeki eğitim görüntülerinden kendi oluşturduğumuz test veri kümelerinin haricinde de web sitesi üzerindeki çevrimiçi test veri kümelerinde ağın performansı değerlendirilmiştir. Elde edilen sonuçları Çizelge 4.9'da verilmiştir.

Çizelge 4.9. KITTI 2012 ve KITTI 2015 çevrimiçi test kümelerinde orijinal giriş metotları ve CR-RNN_P ağ yapısı ile elde edilen (%) hata oranlarının karşılaştırılması (Emlek ve Peker, 2021a)

Mimari	KITTI 2012				KITTI 2015		
	Out-All				All / All		
	>5px	>4px	>3px	>2px	D1-bg	D1-fg	D1-all
MC-CNN-Hızlı	15.77	16.52	17.58	19.72	15.38	28.84	17.62
MC-CNN-Doğru	14.17	14.77	15.65	17.62	13.29	28.29	15.79
MC-CNN-Hızlı + CR-RNNP	13.33	13.83	14.54	15.99	12.98	28.36	15.54
CBCA	31.00	32.52	34.51	37.95	30.81	35.51	31.59
CBCA+ CR-RNNP	26.58	27.97	29.98	33.92	26.56	30.71	27.25

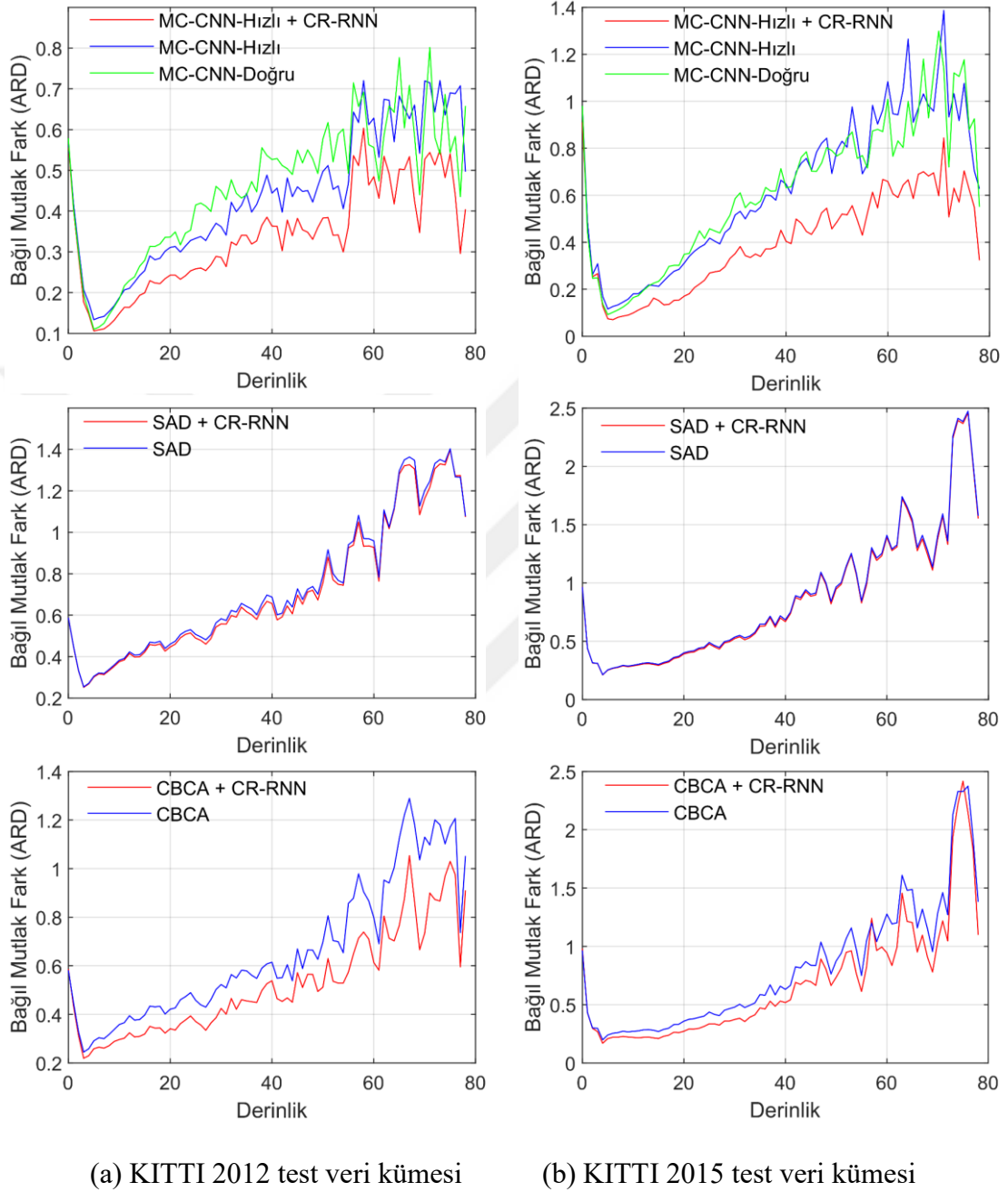
KITTI 2012 ve KITTI 2015 çevrimiçi test veri kümesi üzerindeki test sonuçlarına göre, önerdiğimiz yöntemin hata oranlarında ortalama olarak yaklaşık %15'lik bir iyileştirme sağladığı görülmektedir. Çizelge 4.9'da görüldüğü gibi tüm kötü piksel yüzdesi oranlarında, KITTI 2012 veri setinde önemli oranda iyileştirmeler bulunmaktadır. Benzer şekilde, KITTI 2015 veri kümesinde de arka plan ve genel bölgeler için daha iyi performans elde edilmiştir. Fakat CR-RNN yöntemi ön plandaki sahneler için MC-CNN-Hızlı mimarisinden daha iyi sonuçlar elde ederken, MC-CNN-Doğru mimarinin başarısına yaklaştığı görülmektedir.

CBCA'dan elde edilen eşleşme maliyetlerine CR-RNN_P ağ yapısı uygulandığı durumda ise her iki veri kümesi için doğruluk oranlarının MC-CNN'de olduğu gibi yaklaşık %15 oranında iyileştiği gözlenmiştir.

4.2.5 Nitel karşılaştırmalar

Nitel karşılaştırmalarda hem test veri kümelerindeki ARD eğrilerinden hem de çevrim içi test veri kümeleri üzerindeki eşitsizlik haritalarının görsellerinden faydalanılmaktadır.

Nitel karşılaştırmalar ilk olarak ARD eğrileri üzerinde gerçekleştirilmiştir. Şekil 4.6’da KITTI 2012 ve KITTI 2015 test veri kümelerindeki ARD eğrilerinin karşılaştırılmaları verilmektedir.



Şekil 4.6. KITTI 2012 ve KITTI 2015 test veri kümelerindeki ARD eğrileri (Emlek ve Peker, 2021a)

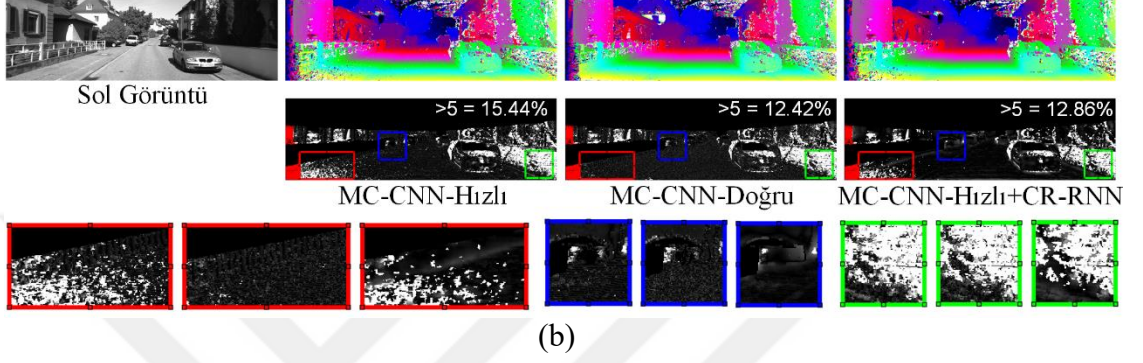
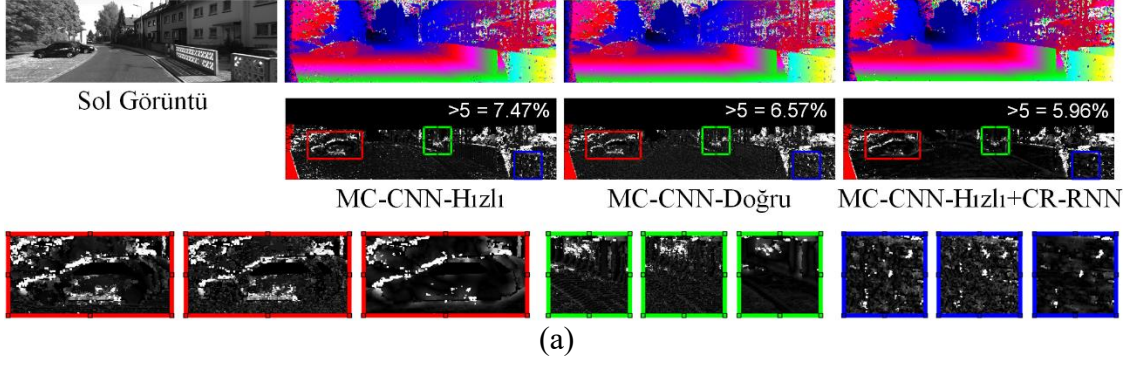
Şekil 4.6’da verilen grafiklerde yatay eksenler derinlik değerini ve dikey eksenler ise ARD değerlerini göstermektedir. Şekil 4.6’da, ilk satırda MC-CNN-Hızlı ve MC-CNN-Doğru mimarisi ile MC-CNN-Hızlı ağından elde edilen eşleşme maliyetlerine CR-RNN ağına en iyi performans gösteren yapısı uygulandığında ortaya çıkan sonuçların

karşılaştırılması verilmektedir. İkinci ve üçüncü satırlardaki karşılaştırmalarda ise sırayla SAD ve CBCA stereo metotlarından elde edilen eşleşme maliyetlerini giriş olarak kullanan CR-RNN ağ yapısının performansları verilmektedir.

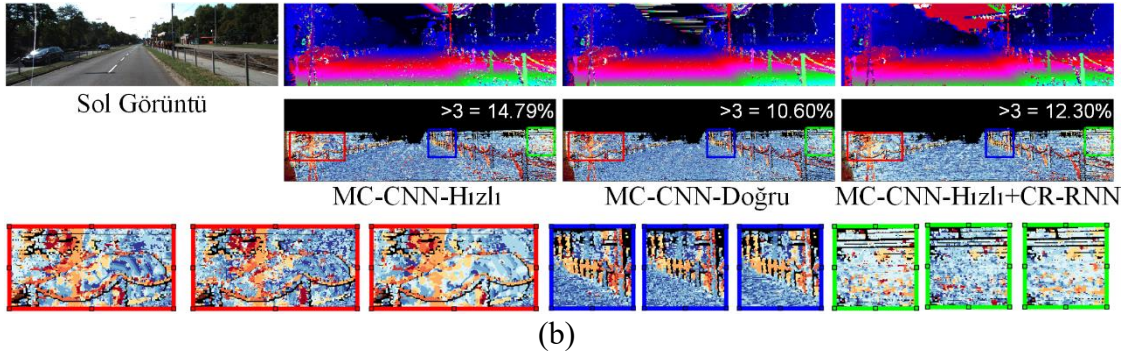
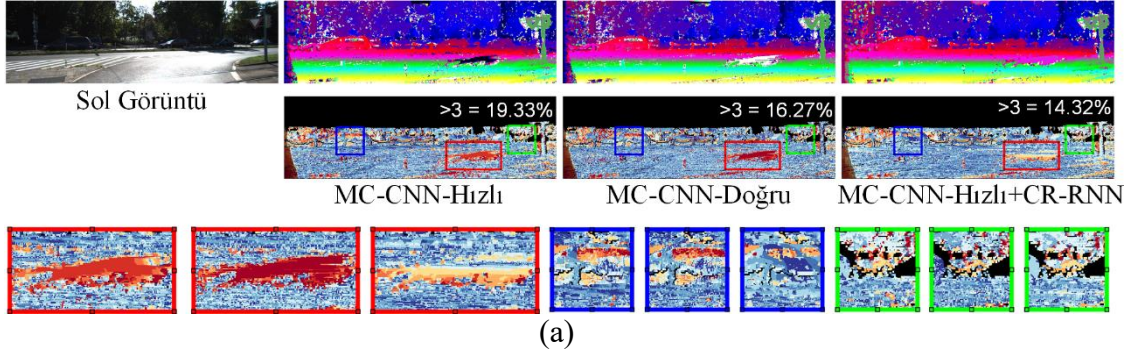
Şekil 4.6’da görüldüğü üzere 3B düzlemdeki derinlik değerlerinde önerilen CR-RNN ağ yapısı diğer modellere göre oldukça düşük hata oranları vermektedir. Aynı zamanda Şekil 4.6’daki grafiklerde derinlik verilerinin sıralı yapıları incelendiğinde CR-RNN ağ yapısının çıkışlarının daha yumuşak geçişlere sahip olduğu gözlemlenmektedir.

Bir diğer nitel karşılaştırma olan KITTI 2012 ve KITTI 2015 çevrimiçi test veri kümeleri üzerindeki derinlik haritalarının karşılaştırmaları için Şekil 4.7, Şekil 4.8, Şekil 4.9 ve Şekil 4.10’da verilen görseller kullanılmaktadır. Şekil 4.7 ve Şekil 4.8’de KITTI 2012 ve KITTI 2015 veri kümelerinde verilen görüntülerde bir tanesi iyi sonucumuzu diğeri ise kötü sonucumuzu temsil etmektedir.

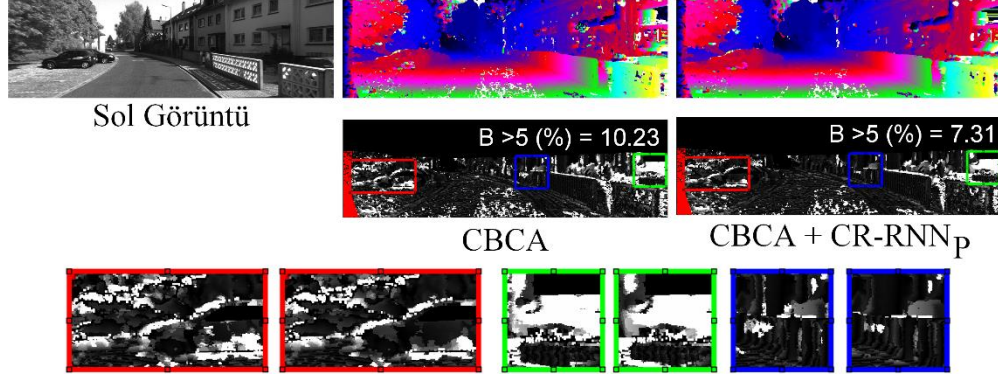
Şekil 4.7, Şekil 4.8, Şekil 4.9 ve Şekil 4.10’da verilen şekillerde yer alan görüntülerde sol ilk satırdaki görüntü giriş görüntüsünü göstermektedir. İlk satırda giriş görüntüsünün yanında ki görüntüler eşitsizlik haritalarını temsil etmektedir. Soldan sağa doğru verilen bu eşitsizlik haritaları sırayla ilgili sütunun altında yazan yöntemlere ait görsellerdir. İkinci satırda yer alan görüntüler bir üst satırda verilen eşitsizlik haritalarına ait hata haritalarını göstermektedir. Üçüncü satırda ise hata haritalarında seçilen bölgelerin yakınlaştırılmış hali eşitsizlik haritalarının veriliş sırasına göre görselleştirilmiştir. Bu alanlar hata haritaları üzerindeki düzeltme işlemlerinin daha iyi görülmesini sağlamaktadır.



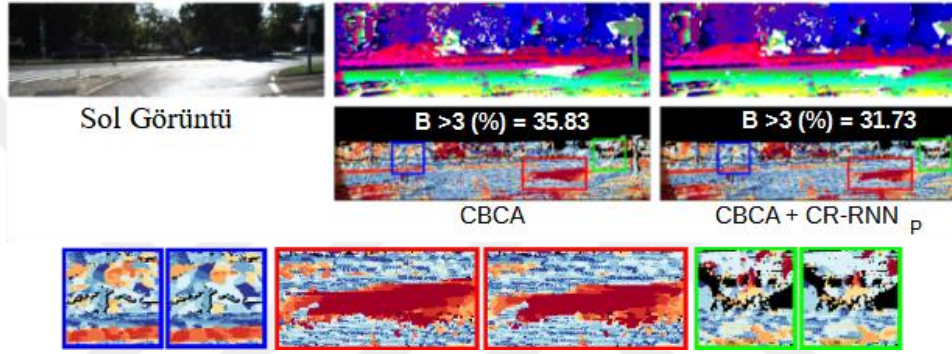
Şekil 4.7 CR-RNN ve MC-CNN yöntemlerinin KITTI 2012 test veri kümesi kullanarak elde ettikleri eşitsizlik ve hata haritaları (Emlek ve Peker, 2021a)



Şekil 4.8. CR-RNN ve MC-CNN yöntemlerinin KITTI 2015 test veri kümesi kullanarak elde ettikleri eşitsizlik ve hata haritaları (Emlek ve Peker, 2021a)



Şekil 4.9. CR-RNN ve CBCA yöntemlerinin KITTI 2012 test veri kümesi kullanarak elde ettikleri eşitsizlik ve hata haritaları (Emlek ve Peker, 2021a)



Şekil 4.10. CR-RNN ve CBCA yöntemlerinin KITTI 2015 test veri kümesi kullanarak elde ettikleri eşitsizlik ve hata haritaları (Emlek ve Peker, 2021a)

Şekil 4.7, Şekil 4.8, Şekil 4.9 ve Şekil 4.10’da verilen eşitsizlik haritaları, CR-RNN ağının MC-CNN ağı ve CBCA yöntemine kıyasla yatay eşitsizlik değerleri arasındaki yumuşak geçişi iyileştirdiğini göstermektedir. Bu yöntemle oluşturulan eşitsizlik haritalarının diğer yöntemlerle elde edilen eşitsizlik haritalarına göre daha az gürültülü bölge içerdiği gözlemlenmektedir. Ayrıca, önerilen yöntemin yol yüzeyleri gibi desensiz bölgelerde daha az gürültülü ve obje sınırları gibi derinlik bilgilerinin ani değişiklikler gösterdiği yerlerde de daha doğru performans gösterdiği hata görüntülerinden açıkça görülebilmektedir. Bu sonuç, RNN’lerin sıralı bilgilerinin kullanılması nedeniyle beklenen bir durumdur.

Hata haritaları dikkate alındığında, CR-RNN ağının daha geniş bir alanda görsel iyileştirmelere sahip olduğu açıktır. Fakat bu ağ yapısı, eşitsizlik haritasındaki çeşitli bölgelerde hataları yerel olarak düzeltilmediğinden, sayısal hata değerlerindeki azalma yüksek değildir.

Sonuç olarak bu bölümde, stereo görüntülerden elde edilen eşleşme maliyetleri iyileştirilerek eşitsizlik haritalarının başarısını artırmak amacıyla CR-RNN önerilmiştir. Bu ağ yapısında piksellerin ve dolayısıyla eşleşme maliyetlerinin sıralı yapısı kullanılarak eşleşme maliyetlerinin düzeltilmesi amaçlanmıştır. Kullanılan sıralı bilgiler ise RNN ağı ile anlamlı hale getirilmiştir.

CR-RNN, literatürde oldukça popüler olarak kullanılan üç farklı metottan elde edilen eşleşme maliyetlerine uygulanmıştır. Nicel karşılaştırmalarda CR-RNN'nin KITTI 2012 ve KITTI 2015 veri kümelerinde üç farklı metotla elde edilen eşleşme maliyetleri üzerinde doğruluğu artırdığı gösterilmiştir.

Kullanılan bu metotlardan derin öğrenme tabanlı eşleşme maliyeti üretme yöntemi olan MC-CNN-Hızlı yöntemine CR-RNN ağı ile görece birkaç parametre eklenerek ağın başarımı MC-CNN-Doğru yöntemine yaklaştırılmıştır. Ayrıca, CR-RNN'nin başarısı, CBCA ve SAD gibi geleneksel yöntemlerle elde edilen eşleşme maliyetleri kullanılarak da gösterilmiştir.

Benzer şekilde, niteliksel sonuçlar eşitsizlik haritalarındaki iyileştirmeleri daha görünür kılmaktadır. Bu sonuçlar, önerilen yöntemin nesne sınırlarında ve aynı zamanda yollar gibi düşük dokulu alanlarda eşitsizlik değerlerini iyileştirmek için kullanılabileceğini kanıtlamaktadır. Önerilen yöntem basit bir yapıya sahiptir ve bu nedenle, çıktı olarak eşleşme maliyetleri veren yöntemlere kolayca uyarlanabilir.

Gelecekteki çalışmalarda, performansı daha da iyileştirmek için uzun kısa süreli bellek (Long Short-Term Memory, LSTM) ve kapılı tekrarlayan birimler (Gated Recurrent Units, GRU) tabanlı modüllerin uygulanması düşünülebilir. Bunların yanı sıra, komşu piksel grupları ağlara giriş olarak eklenmeye çalışılarak ağın başarısı artırılabilir.

BÖLÜM V

YATAY DİKKATLİ EŞİTSİZLİK AĞI

Görüntü işleme veya bilgisayarlı görü alanındaki çalışmalarda ilk adım olarak genellikle öznitelik çıkarma işlemleri uygulanmaktadır. Öznitelik çıkarma işlemleri hem geleneksel hem de derin öğrenme tabanlı yöntemler ile gerçekleştirilebilmektedir. Stereo eşleştirme yöntemlerinde de diğer alanlarda olduğu gibi, ilk adım olarak stereo görüntüler üzerinden öznitelik çıkarma işlemleri gerçekleştirilmektedir.

Literatürdeki derin öğrenme tabanlı yöntemlerden bazıları, öznitelik çıkarma işleminde stereo görüntüleri birleştirip (Mayer vd., 2016) giriş olarak kullanırken, bazı yöntemler ise sol ve sağ görüntüleri ayrı ayrı (Mayer vd., 2016; Kendall vd., 2017; Chang ve Chen, 2018) kullanmaktadır. Stereo görüntüleri ayrı ayrı ele alan bu yöntemler, görüntüler üzerinde siyam ağırları kullanarak öznitelik haritaları üretmektedir. Stereo görüntüleri ayrı ayrı kullanan yöntemlerin, stereo görüntüleri birleştirerek kullanan yöntemlere göre daha başarılı sonuçlar elde ettiği gösterilmiştir. (Mayer vd., 2016).

Stereo eşleştirme işleminde ön işlem olarak stereo kameralardan elde edilen görüntüler düzeltilerek aynı yatay eksen üzerine getirilmektedir. Böylece, sol görüntüdeki her bir pikselin eşleştirilmesi gereken piksel karşılığı sağ görüntüde aynı yatay eksen üzerinde bulunmaktadır. Dolayısıyla stereo eşleştirme işlemlerinde diğer görüntü işleme uygulamalarından farklı özel bir durum ortaya çıkmaktadır. Dolayısıyla diğer görüntü işleme uygulamalarında kullanılan öznitelik çıkarma işlemlerinin, stereo eşleştirme işlemlerinde başarıyı artırabilmesi için uygulama bazlı özelleştirilmesi gerektiği düşünülmüştür.

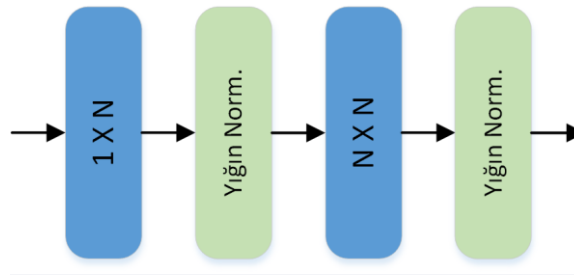
Bu bölümde, yatay eksenindeki piksellerin içerdiği sıralı bilgiler kullanılarak stereo eşleştirme işlemlerindeki öznitelik çıkarma işlemlerinde ekstra bilgi sağlamak amaçlanmıştır. Bu amaç doğrultusunda yatay dikkatli evrişim (Horizontal Attention Convolution - HAC) yapısı önerilmiştir. HAC yapısı, stereo görüntülerdeki yatay düzlemde bulunan sıralı bilgilerin derin öğrenmeye dayalı yöntemlerle daha iyi çıkarılabilmesini sağlamaktadır. Önerilen HAC yapısının öznitelik çıkarma işlemindeki başarısını gözlemlemek için yapı literatürde kabul görmüş olan DispNetC ağının öznitelik

ıkarma katmanına entegre edilmiřtir ve oluřturulan aęa Yatay Dikkatli Eřitsizlik Aęı (Horizontal Attention Disparity Network, HADN) adı verilmiřtir.

Önerilen HADN aęının eęitimi ve test iřlemleri Scene Flow veri kümesindeki görüntüler kullanılarak gerekleřtirilmiřtir. Elde edilen sonular yine bu veri kümesinde yer alan test kümesindeki görüntüler kullanılarak gösterilmiřtir. Test iřlemlerinde yapılan karřılařtırmalar doęrultusunda uygulanan HAC yapısı aęın bařarisını artırmıřtır. Buna ek olarak yapının toplam alıřma süresi üzerindeki etkisinin de ihmal edilecek kadar düşük olduęu gözlemlenmiřtir. Bu bölümde yapılan alıřmalar IEEE kapsamında taranan bir konferansta sunulmuř ve yayınlanmıřtır (Emlek ve Peker, 2021b).

5.1 Aę Yapısı

Standart derin öęrenme tabanlı görüntü iřleme alıřmalarındaki öznitelik ıkarma iřlemlerinde farklı boyutlarda $N \times N$ 'lik ekirdek yapılarına sahip evriřim iřlemleri kullanılmaktadır. Dięer görüntü iřleme uygulamalarından farklı olarak stereo eřleřtirme iřlemlerinde, eřleřtirilmesi gereken pikseller aynı yatay ekseninde bulunmaktadır. Dolayısıyla stereo eřleřtirme iřlemlerinde kullanılan evriřim iřlemlerinin standart ekirdek yapısı deęiřtirilmiřtir. Bu ekirdek yapısında yataydaki sıralı bilgilerin daha iyi ortaya ıkarılabilmesi amacıyla ekirdek yapısındaki yatay öncelikler ve aęırlıklar ön plana ıkarılmıřtır. Öznitelik ıkarma iřleminde yataydaki öznitelikleri daha iyi tanımlayabilmek amacıyla önerilen yapıya yatay dikkatli evriřim (HAC) yapısı adı verilmiřtir. Önerilen HAC yapısı řekil 5.1'de gösterilmiřtir.



řekil 5.1. Yatay dikkatli evriřim (HAC) yapısı (Emlek ve Peker, 2021b)

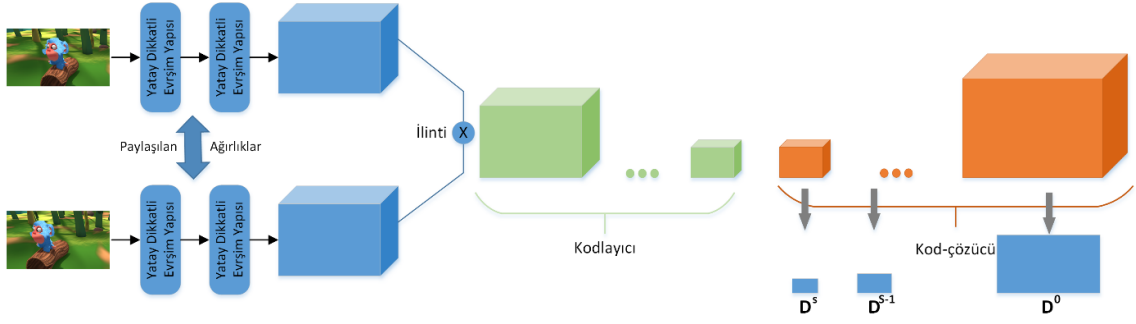
Yatay dikkatli evriřim yapısında ilk olarak $1 \times N$ boyutunda çekirdeęe sahip bir asimetrik evriřim iřlemi uygulanmaktadır. Evriřim iřleminden sonra ise sırayla ReLU aktivasyon fonksiyonu ve yığın normalizasyonu iřlemleri uygulanmaktadır. Daha sonra ise $N \times N$ 'lik bir çekirdeęe sahip evriřim katmanı ile genel özniteliklerin elde edilmesi sağlanmaktadır. Son olarak ise ilk evriřim iřleminde olduęu gibi ReLU aktivasyon fonksiyonu ve yığın normalizasyonu iřlemi uygulanmaktadır.

HAC yapısının stereo eřleřtirme iřleminde ki öznitelik çıkarma iřlemlerinde etkisini gözlemlemek amacıyla DispNetC ağı sečilmiřtir. DispNetC literatürde başarımı ile kabul görmüş gerçek zamanlı uygulamalar için de ideal bir ağ yapısıdır. DispNetC ağ yapısı ile eřleşme maliyetlerinin elde edilmesinde sol ve sağ görüntüden çıkarılan öznitelik haritaları kullanılmaktadır. Öznitelik haritalarını elde etmek için sırayla 7×7 ve 5×5 'lik çekirdeklere sahip iki adet evriřim iřlemi uygulanmaktadır. DispNetC ağında eřsitsizlik haritaları eřleşme maliyetleri üzerinden kodlayıcı-kod çözücü yapısı ile elde edilmiřtir.

DispNetC ağının öznitelik çıkarma katmanında bulunan iki adet evriřim iřleminin yerine önerilen yatay dikkatli evriřim yapısı kullanılmıřtır. Yatay dikkatli evriřim yapısındaki çekirdeklerin boyutlarını belirleyen N deęeri ise DispNetC ağındaki çekirdek boyutları ile aynı tutularak sırayla 7 ve 5 sečilmiřtir.

DispNetC ağının öznitelik çıkarma katmanını yatay dikkatli evriřim yapısı ile modifiye ederek önermiř olduęumuz uçtan uca eřsitsizlik ağ yapısı Şekil 5.2'de verilmektedir. Bu ağ yapısı ise Yatay Dikkatli Eřsitsizlik Ağı (Horizontal Attention Disparity Network - HADN) olarak adlandırılmıřtır.

Şekil 5.2'de verilen modelde kod çözücü iřleminde her bir katmanda eřsitsizlik haritası elde edilmektedir. Farklı katmanlardan elde edilen eřsitsizlik haritaları D ile gösterilmektedir. D ifadesinin üst indisi s ise elde edilen eřsitsizlik haritasının ölçeğini temsil etmektedir. Eęitim iřlemi sırasında elde edilen eřsitsizlik haritaları ölçek oranları ile birleřtirilerek kayıp fonksiyonu elde edilmektedir. Test iřlemleri sırasında ise sadece ağın en son katmanından elde edilen nihai eřsitsizlik haritası (D^0) kullanılmaktadır.



Şekil 5.2. Yatay dikkatli eşitsizlik ağı (HADN) (Emlek ve Peker, 2021b)

5.2 Deneysel Sonuçlar

Yatay Dikkatli Eşitsizlik Ağının eğitim ve test işlemlerinde Scene Flow veri kümesi kullanılmıştır. Eğitim sırasında kayıp değerlerini belirlemek için L1 regresyon fonksiyonu kullanılmıştır. L1 regresyon fonksiyonu Denklem (5.1)'de tanımlanmıştır.

$$loss_D = \frac{1}{N} \sum_{n=1}^N |y_{gercek} - y_{tahmin}| \quad (5.1)$$

Denklem (5.1) 'de verilen kısaltmalardan, y_{gercek} doğru eşitsizlik haritasını, y_{tahmin} ise elde edilen eşitsizlik haritasını göstermektedir. Önerilen Yatay Dikkatli Eşitsizlik ağının kod çözücü yapısındaki her bir katmandan elde edilen eşitsizlik haritası için Denklem (5.1)'de verilen L1 regresyon fonksiyonu kullanılarak kayıp değerleri hesaplanmaktadır. Farklı ölçeklerdeki eşitsizlik haritalarından elde edilen kayıp değerleri Denklem (5.2)'de gösterildiği gibi ölçek oranları ile doğru orantılı olarak birleştirilerek toplam kayıp değeri elde edilmektedir.

$$Total_{kayıp} = \sum_{s=0}^4 \frac{1}{2^s} Loss_D^s \quad (5.2)$$

Denklem (5.2)'de, $Total_{kayıp}$ toplam kayıp değerini, s terimi ise eşitsizlik haritalarının boyut ölçeklerini göstermektedir. s teriminin 0 olduğu durum orijinal görüntü boyunu temsil etmektedir. s teriminde gerçekleşen her bir artış ise görüntü boyutundaki derinlik ve yükseklik değerinin yarıya düştüğünü göstermektedir.

Toplam kayıp deęerini en aza indirmek için ADAM yöntemi kullanılmıştır. Eğitim parametreleri olarak öğrenme oranı 0.001, beta deęerleri ise sırasıyla 0.9 ve 0.999 olarak seçilmiştir. Ağın eğitimi için Scene Flow veri kümesinde 50 tekrar (epoch) gerçekleştirilmiştir.

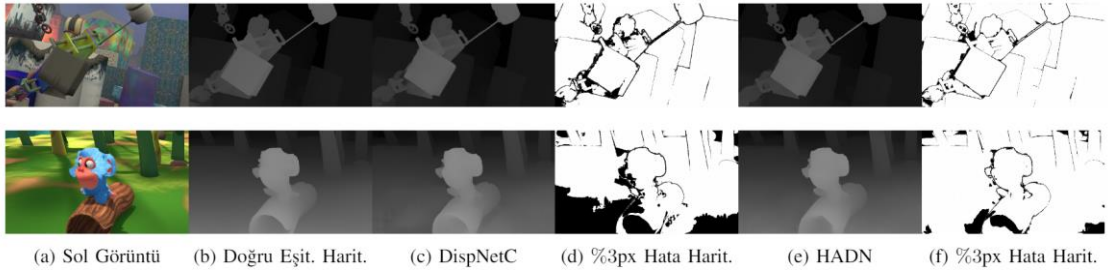
HADN ağının başarısını test etmek için hem nicel hem de nitel karşılaştırmalar yapılmıştır. Nicel karşılaştırmalar için EPE (end point error) hata metrięi kullanılmıştır. HADN ağı, referans olarak alınan DispNetC ağına ilaveten uçtan uca derin öğrenme yöntemi ve eşitsizlik haritaları üreten dięer yöntemlerle karşılaştırılmıştır. Scene Flow veri kümesi kullanılarak elde edilen EPE sonuçları Çizelge 5.1'de verilmiştir.

Çizelge 5.1. HADN, DispNetC ve dięer yöntemlerin Scene Flow veri kümesinde EPE hata oranının ve çalışma sürelerinin karşılaştırılması (Emlek ve Peker, 2021b)

Yöntemler	GC-Net	PSMNet	DispNetC	HADN
EPE	2.51	1.09	1.68	1.46
Süre(saniye)	0.9	0.41	0.06	0.06

Çizelge 5.1'de, HADN ve DispNetC ağ yapıları karşılaştırıldığı zaman HADN ağ yapısındaki EPE hata oranının %13'lük bir azalma sağladığı görülmektedir. Ayrıca HADN ve DispNetC ağ yapılarının çalışma süreleri karşılaştırıldığında eklenen iki adet evrişim işleminin ağıdaki dięer işlemlerin yanında etkisinin ihmal edilebilir düzeyde olduęu gözlemlenmiştir.

Eşitsizlik haritaları üzerinde yapılan nicel karşılaştırmalara ek olarak nitel karşılaştırmalar yapılmıştır. Nitel karşılaştırma işlemi doğrudan referans olarak alınan DispNetC ağı ile gerçekleştirilmiştir. Bu karşılaştırmalar Şekil 5.3'de verilmiştir. Verilen örneklerden ilk satırdaki görüntü FlyingThings3D alt veri kümesine ikinci satırdaki görüntü ise Monkaa alt veri kümesine aittir.



Şekil 5.3. Scene Flow veri kümesi üzerindeki nitel karşılaştırmalar (Emlek ve Peker, 2021b)

Şekil 5.3’de verilen DispNetC ve HADN ağ yapıları ile elde edilen eşitsizlik haritaları karşılaştırıldığında, HADN ağı ile edilen görüntülerde yer alan nesnelerin sınır bölgelerinde yumuşak geçişler yerine daha keskin geçişler olduğu gözlemlenmektedir. Elde edilen hata haritaları karşılaştırıldığında ise, HADN ağı ile elde edilen sonuçlarda çok daha net bir şekilde hata miktarının azlığı dikkat çekmektedir. Özellikle Monkaa alt veri kümesine ait görüntü üzerindeki karşılaştırmalarda HADN ağında, DispNetC ağında hatalı olarak üretilen kameraya yakın yerlerdeki arka plan verilerinin neredeyse tamamının düzeltildiği görülmektedir.

Sonuç olarak bu bölümde, stereo eşleştirme işleminde piksellerin sıralı yapısını kullanmak amacıyla yatay dikkatli evrişim (HAC) katmanı önerilmiştir. Önerilen HAC yapısının literatürde kabul görmüş olan DispNetC ağının öznelilik çıkarma katmanına uyarlanması ile yatay dikkatli eşitsizlik ağı (HADN) elde edilmiştir.

HAC yapısının DispNetC ağı üzerindeki etkisini gözlemlemek için Scene Flow veri kümesi üzerinde nitel ve nicel karşılaştırmalar yapılmıştır. Nicel karşılaştırmalar sonunda HAC yapısının DispNetC ağının başarısını yaklaşık %13 oranında artırdığı gözlemlenmiştir.

HAC yapısı ile DispNetC ağı nitel olarak karşılaştırıldığında ise, HAC yapısının eşitsizlik haritalarının doğruluklarını önemli derecede artırdığı gözlemlenmiştir. Özellikle HAC yapısı ile eşitsizlik haritalarında ki nesnelerin kenar bölgelerinde daha keskin ve doğru sonuçlar elde edilmiştir. Ayrıca bu çalışmada eklenen harici iki evrişim işleminin ise çalışma süresi üzerinde önemli derecede bir etkisi olmamıştır.

HAC yapısı derin öğrenme tabanlı bütün stereo eşleştirme yöntemlerindeki öznitelik çıkarma işlemlerinde basitçe kullanılabilecek bir yapıdır. Bu yapının DispNetC ağı üzerindeki etkisi bu çalışma ile gösterilmiştir.



BÖLÜM VI

PARALEL PİRAMİT HAVUZU TABANLI STEREO AĞI

Eşitsizlik haritalarının doğruluğunu etkileyen en önemli faktörlerden biri eşleşme maliyetleridir. Eşleşme maliyetleri ne kadar doğru, ne kadar güvenilir olursa doğru eşitsizlik değerine ulaşmakta o kadar kolay ve yüksek başarımlı olmaktadır. Eşleşme maliyetleri ise temelde her bir pikseli temsil eden öznitelik haritalarının karşılaştırılması ile elde edilmektedir. Derin öğrenme tabanlı yöntemlerden önceki öznitelik çıkarma işlemleri, ilgili pikselin komşu piksellerle olan ilişkileri üzerine kurulmuştur (Kanade ve Okutomi, 1991; Scharstein ve Szeliski, 2002; K. Zhang vd., 2009). Derin öğrenme tabanlı öznitelik çıkarma işlemindeki ilk çalışmalarda geleneksel yöntemlerden esinlenilmiştir ve öznitelik çıkarma işlemi ilgili piksellerin çevresindeki belirli alanlar kullanılarak gerçekleştirilmiştir. (Luo vd., 2016; Žbontar ve Lecun, 2015). İlerleyen çalışmalarda ise derin öğrenme ile birlikte sadece ilgili piksellerin çevresindeki alan değil aynı zamanda içinde bulunduğu nesne veya görüntü üzerindeki küresel içerik bilgileri de özniteliklere aktarılmaya başlanmıştır (Chang ve Chen 2018; Liang vd., 2018; Guo vd., 2019). Bu yöntemlerde küresel içerik bilgisi ile bölgesel özniteliklerin birleştirilmesi amaçlanmıştır.

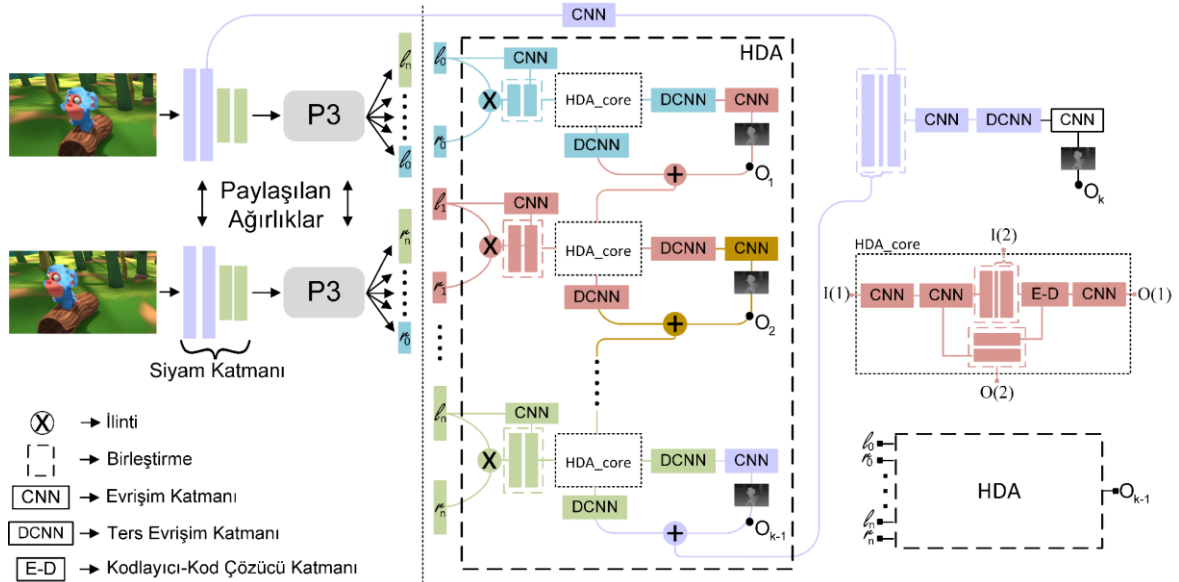
Bu bölümde, küresel ve bölgesel içerik bilgilerini daha kapsamlı ve doğru bir şekilde birleştirebilmek amacıyla el yapımı bir öznitelik çıkarma yöntemi olan ölçek bağımsız öznitelik dönüşümü (Scale-invariant feature transform, SIFT) (Lowe, 2004) yönteminden esinlenerek bir model önerilmektedir. SIFT yöntemi, görüntü işleme ve bilgisayarlı görü alanında öznitelik çıkarma işlemlerinde sıklıkla kullanılan popüler bir öznitelik çıkarma algoritmasıdır. SIFT algoritmasından ilham alınarak derin öğrenme yapılarında öznitelik çıkarma katmanında kullanılabilecek paralel piramit havuzu (parallel pyramid pooling, PPP - P3) yapısı önerilmiştir. Önerilen paralel piramit havuzu kullanılarak da gerçek zamanlı eşitsizlik haritası elde etmek amacıyla bir ağ yapısı önerilmiştir. Önerilen ağ yapısı ise paralel piramit havuzu tabanlı stereo ağı (parallel pyramid pooling stereo network – P3SNet) olarak adlandırılmıştır. Kısaltma ifadesini basitleştirebilmek için ”PPP” yerine “P3” ifadesi tercih edilmiştir.

P3SNet ağıın eğitim ve test işlemleri üç farklı veri kümesinde gerçekleştirilmiştir. Elde edilen sonuçlar literatürdeki öne çıkan yöntemlerle karşılaştırılarak sunulmuştur. P3SNet ağıın

makale çalışmaları tamamlanmış olup değerlendirilmek üzere SCI kapsamındaki bir dergiye gönderilmiştir.

6.1 Ağ Yapısı

Stereo görüntülerden gerçek zamanlı eşitsizlik haritaları elde etmek amacıyla önerilen P3SNet ağ yapısı Şekil 6.1’de verilmektedir. P3SNet ağ yapısında ilk olarak stereo görüntülerden başlangıç öznitelik haritaları elde edilmektedir. Daha sonra bu öznitelik haritaları paralel piramit havuzu olarak adlandırdığımız öznitelik çıkarma yapısına verilmektedir. Paralel piramit havuzu yapısı ile başlangıç özniteliklerinden n farklı boyutta öznitelik haritasının elde edilmesi sağlanmaktadır. Sol ve sağ görüntüden elde edilen farklı boyutlardaki özniteliklerden ilinti işlemi ile farklı boyutlarda n adet eşleşme maliyeti elde edilmektedir. Bu eşleşme maliyetlerinden hiyerarşik eşitsizlik toplama (Hierarchical Disparity Aggregation, HDA) yapısı ile en küçük boyuta sahip olandan başlayarak, kabadan inceye bir eğitim yapısı ile farklı ölçeklerde eşitsizlik haritaları elde edilmektedir. Küçük ölçeklerden elde edilen eşitsizlik haritaları ve öznitelik haritaları üst ölçekteki eşitsizlik haritalarının elde edilmesinde kullanılmaktadır. Bu ağ yapısındaki her bir aşama sıralı başlıklar halinde verilmektedir.



Şekil 6.1. Paralel piramit havuzu tabanlı stereo ağ yapısı

6.1.1 Başlangıç özniteliklerinin çıkarımı

P3SNet ağ yapısındaki öznitelik çıkarma işleminde, ilk olarak giriş olarak alınan stereo görüntüler siyam ağlara verilmektedir. Siyam ağlarda sıralı evrişim işlemleri uygulanan sol ve sağ görüntülerden başlangıç öznitelik haritaları elde edilmektedir. Evrişim işlemlerinin sonunda elde edilen bu öznitelik haritalarının genişlik ve yükseklik boyutları giriş görüntülerinin dörtte biri kadardır. Bu katmanda elde edilen öznitelikler bölgesel içerik bilgilerini çıkarmayı amaçlamaktadır. Başlangıç öznitelik haritaları küresel içerik bilgilerinin elde edilmesi amacıyla bir sonraki bölümde detaylı olarak bahsedilecek olan paralel piramit havuzu yapısına giriş olarak verilmektedir. Başlangıç özniteliklerinin çıkarılmasında uygulanan işlemler Çizelge 6.1’de ayrıntılı olarak verilmiştir.

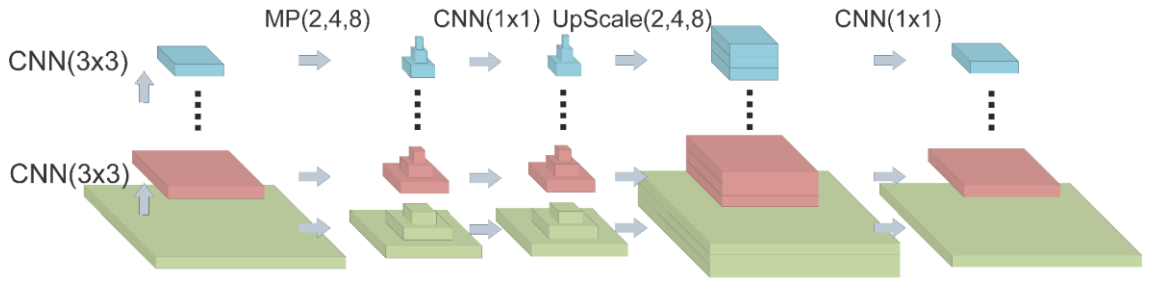
Çizelge 6.1. Başlangıç özniteliklerinin çıkarılmasında kullanılan alt ağ yapısına ait katman giriş/çıkış topolojisi

Katman İsmi	Katman Girişi	Çekirdek	Giriş Boyutu	Çıkış Boyutu
f_1	I_{sol}, I_{sag}	$7 \times 7, 2 + BN + ReLU$	$256 \times 512 \times 3$	$128 \times 256 \times 32$
f_2	f_1	$3 \times 3, 1 + BN + ReLU$	$128 \times 256 \times 32$	$128 \times 256 \times 32$
f_3	f_2	$5 \times 5, 2 + BN + ReLU$	$128 \times 256 \times 32$	$64 \times 128 \times 32$
f_4	f_3	$3 \times 3, 1 + BN + ReLU$	$64 \times 128 \times 32$	$64 \times 128 \times 32$
f_5	f_4	$3 \times 3, 1 + BN + ReLU$	$64 \times 128 \times 32$	$64 \times 128 \times 32$
f_6	f_5	$3 \times 3, 1 + BN + ReLU$	$64 \times 128 \times 32$	$64 \times 128 \times 32$
f_7	f_6	$3 \times 3, 1 + BN + ReLU$	$64 \times 128 \times 32$	$64 \times 128 \times 32$
f_8	f_7	$3 \times 3, 1 + BN + ReLU$	$64 \times 128 \times 32$	$64 \times 128 \times 32$

Çizelge 6.1’de siyam ağlardaki her bir katmana giren verilerin isimleri, giriş boyutları, çıkış boyutları ve ilgili katmanda uygulanan işlemler verilmektedir. Bu tabloda verilen kısaltmalarda I_{sol} ve I_{sag} sırasıyla sol ve sağ kameradan elde edilen düzeltilmiş görüntüleri, BN (batch normalization) ise yığın normalizasyonu işlemini temsil etmektedir.

6.1.2 Paralel piramit havuzu yapısı

Paralel piramit havuzu yapısı, giriş olarak aldığı başlangıç özniteliklerinden yerel ve küresel içerik bilgilerini çıkararak birleştirmeyi amaçlamaktadır. Önerilen paralel piramit havuzu yapısı Şekil 6.2’de verilmektedir.



Şekil 6.2. Paralel piramit havuzu yapısı

Şekil 6.2’de görüldüğü üzere, alınan başlangıç öznetelik haritalarına sıralı olarak $n - 1$ aşağı örneklemeli evrişim işlemi uygulanmaktadır. Aşağı ölçekleme işlemi sonrasında n tane farklı ölçekte öznetelik haritası elde edilmektedir. Daha sonra her biri farklı ölçekteki öznetelikler piramit havuzu yapısına giriş olarak verilmektedir.

n farklı ölçekteki özneteliklerin hepsi ayrı ayrı üç farklı boyutta en büyük havuzlama (max-pooling, MP) işlemine tabi tutulmaktadır. Bu işlemler ile her bir ölçeğin kendi içinde farklı ölçekteki öznetelikleri oluşturulmaktadır. Farklı ölçekteki bu özneteliklere 1×1 ’lik evrişim işlemleri uygulanmaktadır. Uygulanan evrişim işlemleri ile elde edilen öznetelikler havuzlama yapısına girmeden önceki boyutlarına çift doğrusal aradeğerleme yöntemi ile ölçeklenmektedir. Daha sonra bu ölçeklenmiş öznetelikler birleştirilmektedir. Her bir ölçekteki nihai öznetelik haritası ise birleştirilmiş özneteliklere 1×1 ’lik bir evrişim işlemi uygulanmasının sonunda elde edilmektedir. Üç katmanlı ($n = 3$) bir yapı için paralel piramit havuzu yapısında uygulanan işlemler Çizelge 6.2’de verilmiştir.

Önerilen paralel piramit havuzu yapısı PSMNet ağında kullanılan havuzlama yapısına göre dört temel farklılık barındırmaktadır. Bu farklılardan ilki mimarilerde bulunan havuzlama işlemlerinde uygulanan yapıların farklılıklarıdır. PSMNet mimarisinde farklı boyutlarda ortalama havuzlama (Average-Pooling, AP) işlemi uygulanırken, önerilen yapıda ise en büyük havuzlama (Max-Pooling, MP) işlemi uygulanmaktadır. Bu yaklaşım, baskın özneteliklerin korunmasını sağlamaktadır. İkinci farkı havuzlama işlemlerinde uygulanan kaydırma adımlarının uzunlukları oluşturmaktadır. PSMNet ağında kaydırma uzunlukları 16, 32 ve 64 olarak belirlenmişken önerilen ağ yapısında bu uzunluklar 2, 4 ve 8 olarak belirlenmiştir. Üçüncü farkı önerilen paralel piramit havuzunun genel yapısı oluşturmaktadır. PSMNet ağındaki havuzlama mimarisi tek bir ölçekte giriş ve çıkışa sahip iken önerilen P3 yapısı tek bir ölçekte giriş ve farklı ölçeklerde çıkışlara sahiptir.

Çizelge 6.2. Üç katmalı paralel piramit havuzu yapısında kullanılan alt ağ yapısına ait katman giriş/çıkış topolojisi

Katman İsmi	Katman Girişi	Çekirdek	Giriş Boyutu	Çıkış Boyutu
$f_{8d2}1$	f_8	$3 \times 3,2 + BN + ReLU$	$64 \times 128 \times 32$	$32 \times 64 \times 32$
$f_{8d2}2$	$f_{8d2}1$	$3 \times 3,1 + BN + ReLU$	$32 \times 64 \times 32$	$32 \times 64 \times 32$
$f_{8d3}1$	$f_{8d2}2$	$3 \times 3,2 + BN + ReLU$	$32 \times 64 \times 32$	$16 \times 32 \times 32$
$f_{8d3}2$	$f_{8d3}1$	$3 \times 3,1 + BN + ReLU$	$16 \times 32 \times 32$	$16 \times 32 \times 32$
$PPP3_{mp2}$	$f_{8d3}2$	$MP, 2 \times 2,2$	$16 \times 32 \times 32$	$8 \times 16 \times 32$
$PPP3_{mp4}$	$f_{\{8d3\}}2$	$MP, 4 \times 4,4$	$16 \times 32 \times 32$	$4 \times 8 \times 32$
$PPP3_{mp8}$	$f_{\{8d3\}}2$	$MP, 8 \times 8,8$	$16 \times 32 \times 32$	$2 \times 4 \times 32$
$PPP3_{mp2}^C$	$PPP3_{mp2}$	$1 \times 1,1 + BN + ReLU$	$8 \times 16 \times 32$	$8 \times 16 \times 32$
$PPP3_{mp4}^C$	$PPP3_{mp4}$	$1 \times 1,1 + BN + ReLU$	$4 \times 8 \times 32$	$4 \times 8 \times 32$
$PPP3_{mp8}^C$	$PPP3_{mp8}$	$1 \times 1,1 + BN + ReLU$	$2 \times 4 \times 32$	$2 \times 4 \times 32$
$PPP3_{mp2}^u$	$PPP3_{mp2}^C$	$Upsample, BI$	$8 \times 16 \times 32$	$16 \times 32 \times 32$
$PPP3_{mp4}^u$	$PPP3_{mp4}^C$	$Upsample, BI$	$4 \times 8 \times 32$	$16 \times 32 \times 32$
$PPP3_{mp8}^u$	$PPP3_{mp8}^C$	$Upsample, BI$	$2 \times 4 \times 32$	$16 \times 32 \times 32$
$PPP3_{out}$	$PPP3_{mp2}^u + PPP3_{mp4}^u + PPP3_{mp8}^u$	$1 \times 1,1 + BN + ReLU$	$16 \times 32 \times 96$	$16 \times 32 \times 32$
$PPP2_{mp2}$	$f_{8d2}2$	$MP, 2 \times 2,2$	$32 \times 64 \times 32$	$16 \times 32 \times 32$
$PPP2_{mp4}$	$f_{8d2}2$	$MP, 4 \times 4,4$	$32 \times 64 \times 32$	$8 \times 16 \times 32$
$PPP2_{mp8}$	$f_{8d2}2$	$MP, 8 \times 8,8$	$32 \times 64 \times 32$	$4 \times 8 \times 32$
$PPP2_{mp2}^C$	$PPP2_{mp2}$	$1 \times 1,1 + BN + ReLU$	$16 \times 32 \times 32$	$16 \times 32 \times 32$
$PPP2_{mp4}^C$	$PPP2_{mp4}$	$1 \times 1,1 + BN + ReLU$	$8 \times 16 \times 32$	$8 \times 16 \times 32$
$PPP2_{mp8}^C$	$PPP2_{mp8}$	$1 \times 1,1 + BN + ReLU$	$4 \times 8 \times 32$	$4 \times 8 \times 32$
$PPP2_{mp2}^u$	$PPP2_{mp2}^C$	$Upsample, BI$	$16 \times 32 \times 32$	$16 \times 32 \times 32$
$PPP2_{mp4}^u$	$PPP2_{mp4}^C$	$Upsample, BI$	$8 \times 16 \times 32$	$32 \times 64 \times 32$
$PPP2_{mp8}^u$	$PPP2_{mp8}^C$	$Upsample, BI$	$4 \times 8 \times 32$	$32 \times 64 \times 32$
$PPP2_{out}$	$PPP2_{mp2}^u + PPP2_{mp4}^u + PPP2_{mp8}^u$	$1 \times 1,1 + BN + ReLU$	$32 \times 64 \times 96$	$32 \times 64 \times 32$
$PPP1_{mp2}$	f_8	$MP, 2 \times 2,2$	$64 \times 128 \times 32$	$32 \times 64 \times 32$
$PPP1_{mp4}$	f_8	$MP, 4 \times 4,4$	$64 \times 128 \times 32$	$16 \times 32 \times 32$
$PPP1_{mp8}$	f_8	$MP, 8 \times 8,8$	$64 \times 128 \times 32$	$8 \times 16 \times 32$
$PPP1_{mp2}^C$	$PPP1_{mp2}$	$1 \times 1,1 + BN + ReLU$	$32 \times 64 \times 32$	$32 \times 64 \times 32$
$PPP1_{mp4}^C$	$PPP1_{mp4}$	$1 \times 1,1 + BN + ReLU$	$16 \times 32 \times 32$	$16 \times 32 \times 32$
$PPP1_{mp8}^C$	$PPP1_{mp8}$	$1 \times 1,1 + BN + ReLU$	$8 \times 16 \times 32$	$8 \times 16 \times 32$
$PPP1_{mp2}^u$	$PPP1_{mp2}^C$	$Upsample, BI$	$32 \times 64 \times 32$	$64 \times 128 \times 32$
$PPP1_{mp4}^u$	$PPP1_{mp4}^C$	$Upsample, BI$	$16 \times 32 \times 32$	$64 \times 128 \times 32$
$PPP1_{mp8}^u$	$PPP1_{mp8}^C$	$Upsample, BI$	$8 \times 16 \times 32$	$64 \times 128 \times 32$
$PPP1_{out}$	$PPP1_{mp2}^u + PPP1_{mp4}^u + PPP1_{mp8}^u$	$1 \times 1,1 + BN + ReLU$	$64 \times 128 \times 96$	$64 \times 128 \times 32$

Yapılar arasındaki en son fark ise, SPP yapısının tek çıkışı yukarı örnekleme ve birleştirme işlemleriyle oluşturulurken, P3 yapısının n adet çıkışı çift doğrusal aradeğerleme ve 1×1 evrişim işlemleri ile oluşturulmaktadır.

6.1.3 Eşleşme maliyetleri

Eşleşme maliyetlerinin hesaplanmasında sol ve sağ görüntüden paralel piramit havuzu yapısıyla elde edilen farklı ölçekteki öznitelik haritaları kullanılmaktadır. Paralel piramit havuzu yapısı ile öznitelik çıkarma işleminde olduğu gibi n farklı ölçekte eşleşme maliyetleri elde edilmektedir. Eşleşme maliyetleri elde edilirken her bir ölçekte korelasyon işlemi uygulanmaktadır. Her bir ölçekteki korelasyon işlemi Denklem (6.1)'de tanımlanmaktadır.

$$C_{kor}^{\text{ölçek}}(d, x, y) = \frac{1}{N_c} \langle f_l^{\text{ölçek}}(x, y), f_r^{\text{ölçek}}(x - d, y) \rangle \quad (6.1)$$

Denklem (6.1)'de verilen terimlerden $f_l^{\text{ölçek}}$ ve $f_r^{\text{ölçek}}$ sırasıyla sol ve sağ görüntüden paralel piramit havuzunda belirli ölçeklerden elde edilmiş öznitelik haritalarını temsil etmektedir. d ifadesi ise eşitsizlik değeridir ve sağ görüntüdeki ötelemeyi temsil etmektedir. d değeri özniteliklerinin boyutunun giriş görüntüsünün boyutuna oranı ile doğru orantılı olarak her bir ölçekte değişmektedir. Paralel piramit havuzundan elde edilen n adet öznitelik için gerçekleştirilen n farklı korelasyon işlemlerine ait detaylar Çizelge 6.3'de verilmektedir.

Çizelge 6.3. Paralel piramit havuzundan elde edilen özniteliklere uygulanan korelasyon işlemleri

Katman İsmi	Katman Girişi	İşlem	Giriş Boyutu	Çıkış Boyutu
C_{kor}^n	$Sol_PPPn_{out}, Sağ_PPPn_{out}$	<i>korelasyon</i>	$\frac{64}{2^{(n-1)}} \times \frac{128}{2^{(n-1)}} \times 32,$ $\frac{64}{2^{(n-1)}} \times \frac{128}{2^{(n-1)}} \times 32$	$\frac{64}{2^{(n-1)}} \times \frac{128}{2^{(n-1)}} \times \frac{192}{2^{(n-1)}}$
C_{kor}^3	$Sol_PPP3_{out}, Sağ_PPP3_{out}$	<i>korelasyon</i>	$16 \times 32 \times 32, 16 \times 32 \times 32$	$16 \times 32 \times 32$
C_{kor}^2	$Sol_PPP2_{out}, Sağ_PPP2_{out}$	<i>korelasyon</i>	$32 \times 64 \times 32, 32 \times 64 \times 32$	$32 \times 64 \times 24$
C_{kor}^1	$Sol_PPP1_{out}, Sağ_PPP1_{out}$	<i>korelasyon</i>	$64 \times 128 \times 32, 64 \times 128 \times 32$	$64 \times 128 \times 48$

Eşleşme maliyetlerinin elde edilmesinde kullanılacak yöntem belirlenirken gerçek zamanlı bir ağ yapısı oluşturma düşüncesi ön planda tutulmuştur. Elde edilecek eşleşme maliyetlerinin boyutunun sonraki katmanlarda kullanılacak olan evrişim işlemlerinin de boyutunu etkileyeceği göz önünde bulundurulmuştur. Dolayısıyla dört boyutlu (4B) eşleşme maliyeti elde eden işlemler daha yüksek doğruluk ve temsil yeteneklerine sahip olmalarına rağmen tercih edilmemişlerdir.

6.1.4 Eşitsizlik haritasının elde edilmesi

Paralel piramit havuzundan n farklı ölçekte elde edilen öznitelik haritaları ve bu öznitelik haritalarından evrişim işlemleri sonucu elde edilen eşleşme maliyetleri kullanılarak eşitsizlik haritaları hesaplanmaktadır.

Eşleşme maliyetlerini ve görüntülerden elde edilen öznitelikleri kullanarak eşitsizlik haritalarının elde edilmesi için hiyerarşik eşitsizlik toplama (Hierarchical Disparity Aggregation, HDA) yapısı önerilmiştir. Önerilen HDA yapısında nihai eşitsizlik haritasını elde etmek için kabadan inceye doğru bir eğitim gerçekleştirilmiştir. Dolayısıyla ilk olarak en küçük ölçekteki eşitsizlik haritası elde edilmiştir ve daha sonra kabadan inceye bu eğitim yapısı ile ağın ilk olarak giriş verisinin genel özniteliklerini öğrenmesi ve daha sonra her bir ölçek artırımı ile birlikte görüntüye ait detayları daha iyi öğrenmesi amaçlanmıştır.

HDA yapısının $n \times 2$ adet girişi ve bir adet çıkışı bulunmaktadır. n ağın ölçek boyutudur ve P3 yapısındaki paralel katman sayısını belirtmektedir. HDA yapısı Şekil 6.1’de görüldüğü üzere, giriş olarak eşleşme maliyetlerini ve sol görüntüden P3 yapısıyla elde edilen özniteliklerin yeniden boyutlandırılması ile oluşturulan öznitelikleri almaktadır. Bu girdiler Şekil 6.1’de gösterildiği gibi ölçek değerlerine göre farklı HDA çekirdeklerine (HDA_core) giriş olarak verilmektedir. HDA_core olarak adlandırılan bu alt işlem birimleri ise iki adet girişe ($I(1)$ ve $I(2)$) ve iki adet çıkışa ($O(1)$ ve $O(2)$) sahiptir. Bu girişlerden ilkinde ($I(1)$) eşleşme maliyetlerinden elde edilen birleşik öznitelikler verilmektedir. İkinci girişe ($I(2)$) ise alt ölçekten elde edilen öznitelikler verilmektedir. En düşük ölçekte, ikinci girişten herhangi bir veri girişi olmamaktadır.

HDA_core’da, ilk girişten ($I(1)$) gelen birleşik maliyetlere, birbiri ardına iki konvolüsyon işlemi uygulanmaktadır. Daha sonra elde edilen öznitelikler, ikinci girişten ($I(2)$) alınan alt ölçekten gelen büyütülmüş öznitelikler ile birleştirilmektedir. Daha sonra birleştirilmiş özellikler kodlayıcı-kod çözücü yapısına verilmektedir. Ardından elde edilen özniteliklere bir adet evrişim işlemi daha uygulandıktan sonra HDA_core’un ilk çıkışı olan ($O(1)$) elde edilmektedir. Yapının ikinci çıkışı olan ($O(2)$) ise kodlayıcı-kod çözücü yapısından elde edilen özniteliklerle $I(1)$ ’den alındıktan sonra iki adet evrişim işlemi uygulanan özniteliklerin birleştirilmesi ile elde edilmektedir.

Alt ölçeklerden elde edilen eşitsizlik ve öznitelik haritaları bir sonraki katman için referans olarak kullanılmaktadır. Bu verilerin bir üst katmanda kullanılabilmesi için ölçek değerlerinin yatay ve dikey ekseninde iki katına çıkması gerekmektedir. HDA_core modülünün çıktıları olan öznitelik haritalarının üst ölçeğe yükseltilmesi için ters evrişim işlemleri kullanılmaktadır. Ayrıca HDA yapısında her ölçekte referans bilgi olarak eşitsizlik haritaları üretilmektedir. Eşitsizlik haritaları, HDA_core'ların birinci çıkışı olan $O(1)$ 'lerin yükseltilmiş özniteliklerine evrişim işlemleri uygulanması ile elde edilmektedir.

Nihai eşitsizlik haritasının elde edilmesinde son adım olarak ise öznitelik çıkarma işlemlerinden elde edilen en büyük ölçekli sol görüntünün öz nitelikleri evrişim işlemi ile yeniden boyutlandırılarak HDA'dan elde edilen öznitelikler ile birleştirilir. Daha sonra, orijinal görüntü boyutunu getirmek için tekrar ters evrişim işlemi uygulanır. En son adımda ise evrişim işlemi yardımıyla bu maliyetlerden nihai eşitsizlik haritası elde edilir.

Ağ yapısında uygulanan evrişim işlemlerinde eşitsizlik haritası elde etmek için olanlar hariç yığın normalizasyonu ve ReLU aktivasyon fonksiyonu kullanılmıştır. Üç katmanlı bir P3 yapısı için eşitsizlik haritalarının elde edilmesinde uygulanan işlemler Çizelge 6.4'de verilmektedir.

6.1.5 Kayıp fonksiyonu

Şekil 6.1'de verilen paralel piramit havuzu tabanlı stereo ağ yapısında O_i ($i=1,2,...,k$) ile gösterilen farklı ölçeklere ait her bir eşitsizlik haritası için kayıp değerleri hesaplanmaktadır. Bu kayıp değerleri belirli oranlarda birleştirilerek nihai bir kayıp değeri elde edilmektedir. Ağın eğitiminde ise yumuşatılmış L1 regresyon kayıp fonksiyonu kullanılmaktadır. Yumuşatılmış L1 regresyon kayıp fonksiyonu Denklem (6.2)'de verilmektedir.

Çizelge 6.4. Eşitsizlik haritalarının elde edilmesinde kullanılan alt ağ yapısına ait katman giriş/çıkış topolojisi

Katman İsmi	Katman Girişi	Çekirdek	Giriş Boyutu	Çıkış Boyutu
$Sol_PPP3_{out}rd$	Sol_PPP3_{out}	$3 \times 3,1 + BN + ReLU$	$16 \times 32 \times 32$	$16 \times 32 \times 64$
$C3_1$	$Sol_PPP3_{out}rd, C_{kor}^3$	Bağlama	$16 \times 32 \times 64, 16 \times 32 \times 12$	$16 \times 32 \times 76$
$C3_2$	$C3_1$	$3 \times 3,1 + BN + ReLU$	$16 \times 32 \times 76$	$16 \times 32 \times 256$
$C3_3$	$C3_2$	$3 \times 3,1 + BN + ReLU$	$16 \times 32 \times 256$	$16 \times 32 \times 256$
$C3_4$	$C3_3$	$3 \times 3,2 + BN + ReLU$	$16 \times 32 \times 256$	$8 \times 16 \times 512$
$C3_5$	$C3_4$	$4 \times 4,2 + BN + ReLU$	$8 \times 16 \times 512$	$16 \times 32 \times 256$
$C3$	$C3_3, C3_5$	Bağlama	$16 \times 32 \times 256, 16 \times 32 \times 256$	$16 \times 32 \times 512$
$C3_{DCNN}$	$C3$	$4 \times 4,2 + BN + ReLU$	$16 \times 32 \times 512$	$32 \times 64 \times 128$
F_d_3	$C3_5$	$4 \times 4,2 + BN + ReLU$	$16 \times 32 \times 256$	$32 \times 64 \times 128$
$d3$	F_d_3	$3 \times 3,1$	$32 \times 64 \times 128$	$32 \times 64 \times 1$
$Sol_PPP2_{out}rd$	Sol_PPP2_{out}	$3 \times 3,1 + BN + ReLU$	$32 \times 64 \times 32$	$32 \times 64 \times 64$
$C2_1$	$Sol_PPP2_{out}rd, C_{kor}^2$	Bağlama	$32 \times 64 \times 64, 16 \times 32 \times 24$	$32 \times 64 \times 88$
$C2_2$	$C2_1$	$3 \times 3,1 + BN + ReLU$	$32 \times 64 \times 88$	$32 \times 64 \times 128$
$C2_3$	$C2_2$	$3 \times 3,1 + BN + ReLU$	$32 \times 64 \times 128$	$32 \times 64 \times 128$
$C2_3c$	$C3_{DCNN}, d3, C2_3$	Bağlama	$16 \times 32 \times 128, 16 \times 32 \times 1, 32 \times 64 \times 128$	$32 \times 64 \times 257$
$C2_4$	$C2_3c$	$3 \times 3,2 + BN + ReLU$	$32 \times 64 \times 257$	$16 \times 32 \times 256$
$C2_5$	$C2_4$	$4 \times 4,2 + BN + ReLU$	$16 \times 32 \times 256$	$32 \times 64 \times 128$
$C2$	$C2_3, C2_5$	Bağlama	$32 \times 64 \times 128, 32 \times 64 \times 128$	$32 \times 64 \times 256$
$C2_{DCNN}$	$C2$	$4 \times 4,2 + BN + ReLU$	$32 \times 64 \times 256$	$64 \times 128 \times 64$
F_d_2	$C2_5$	$4 \times 4,2 + BN + ReLU$	$32 \times 64 \times 128$	$64 \times 128 \times 64$
$d2$	F_d_2	$3 \times 3,1$	$64 \times 128 \times 64$	$64 \times 128 \times 1$
$Sol_PPP1_{out}rd$	Sol_PPP1_{out}	$3 \times 3,1 + BN + ReLU$	$64 \times 128 \times 32$	$64 \times 128 \times 64$
$C1_1$	$Sol_PPP1_{out}rd, C_{kor}^1$	Bağlama	$64 \times 128 \times 64, 64 \times 128 \times 48$	$64 \times 128 \times 112$
$C1_2$	$C1_1$	$3 \times 3,1 + BN + ReLU$	$64 \times 128 \times 112$	$64 \times 128 \times 64$
$C1_3$	$C1_2$	$3 \times 3,1 + BN + ReLU$	$64 \times 128 \times 64$	$64 \times 128 \times 64$
$C1_3c$	$C2_{DCNN}, d2, C1_3$	Bağlama	$64 \times 128 \times 64, 64 \times 128 \times 1, 64 \times 128 \times 64$	$64 \times 128 \times 129$
$C1_4$	$C1_3c$	$4 \times 4,2 + BN + ReLU$	$64 \times 128 \times 129$	$128 \times 256 \times 32$
$C1_5$	$C1_4$	$3 \times 3,2 + BN + ReLU$	$128 \times 256 \times 32$	$64 \times 128 \times 64$
$C1$	$C1_3, C1_5$	Bağlama	$64 \times 128 \times 64, 64 \times 128 \times 64$	$64 \times 128 \times 128$
$C1_{DCNN}$	$C1$	$4 \times 4,2 + BN + ReLU$	$64 \times 128 \times 128$	$128 \times 256 \times 32$
F_d_1	$C1_5$	$4 \times 4,2 + BN + ReLU$	$64 \times 128 \times 64$	$128 \times 256 \times 32$
$d1$	F_d_1	$3 \times 3,1$	$128 \times 256 \times 32$	$128 \times 256 \times 1$
$f2_rd$	$f2$	$3 \times 3,1 + BN + ReLU$	$128 \times 256 \times 32$	$128 \times 256 \times 32$
$C0$	$C1_{DCNN}, d1, f2_rd$	Bağlama	$128 \times 256 \times 32, 128 \times 256 \times 1, 128 \times 256 \times 32$	$128 \times 256 \times 65$
$C0_1$	$C0$	$3 \times 3,1 + BN + ReLU$	$128 \times 256 \times 65$	$128 \times 256 \times 32$
F_d_0	$C0_1$	$4 \times 4,2 + BN + ReLU$	$128 \times 256 \times 32$	$256 \times 512 \times 16$
$d0$	F_d_0	$3 \times 3,1$	$256 \times 512 \times 16$	$256 \times 512 \times 1$

$$O_{loss} = \frac{1}{N} \sum_{n=1}^N smooth_{L1}(|Y_{gerçek} - Y_{tahmin}|) \quad (6.2)$$

Denklem (6.2)'de verilen ifadedeki $Y_{gerçek}$ terimi doğru eşitsizlik değerini, Y_{tahmin} ise elde edilen eşitsizlik haritasını temsil etmektedir. $smooth_{L1}$ fonksiyonu ise L1 regresyonunun yumuşatılmış hali olup Denklem (6.3)'de verilmektedir.

$$smooth_{L1}(x) = \begin{cases} 0.5x^2 & x < 1 \\ x - 0.5 & x \geq 1 \end{cases} \quad (6.3)$$

Farklı boyuttaki eşitsizlik haritalarından elde edilen kayıp değerleri ise Denklem (6.4)'de belirtildiği gibi birleştirilmektedir.

$$L_{toplam} = \sum_{s=0}^k \lambda_s O_{loss}^s \quad (6.4)$$

Denklem (6.4)'de verilen s terimi farklı ölçekleri, k terimi P3'de oluşturulan maksimum ölçek değerini, λ değeri ise birleştirme ağırlığını temsil etmektedir ve Denklem (6.5)'de belirtildiği gibi hesaplanmaktadır.

$$\lambda_x = \frac{x}{\sum_{i=0}^k i} \quad (6.5)$$

6.2 Deneysel Sonuçlar

P3SNet ağının eğitim ve test işlemlerinde Scene Flow ve KITTI veri kümeleri kullanılmıştır. Başlangıç eğitimi için Scene Flow veri kümesi kullanılmıştır. Daha sonra KITTI 2012 ve KITTI 2015 eğitim veri kümeleri ile ince ayar işlemi yapılmıştır. Test işlemleri ise Scene Flow veri kümesinde çevrimdışı test görüntüleri kullanılarak, KITTI 2012 ve KITTI 2015 veri kümelerinde ise çevrimiçi test görüntüleri kullanılarak gerçekleştirilmiştir.

Kayıp fonksiyonunu en aza indirmek için ADAM yöntemi kullanılmıştır. Eğitim parametrelerinden öğrenme oranı 0.01, beta değerleri ise sırasıyla 0.9 ve 0.999 olarak seçilmiştir. Ağın başlangıç eğitimi için Scene Flow veri kümesinde 200 tekrar (epoch) gerçekleştirilmektedir. İnce ayar işlemleri ise her iki KITTI veri kümesinde ayrı ayrı 700 tekrar gerçekleştirilmiştir.

Öğrenme oranları bütün veri kümelerindeki eğitimlerde iterasyon sayısına göre güncellenmiştir. Scene Flow veri kümesinde 50 tekrardan sonraki her 50. tekrarda 2'ye bölünerek, KITTI veri kümelerinde ise 200 tekrardan sonra her 75. tekrarda 2'ye bölünerek ayarlanmıştır.

Ağın başarımı hem nitel hem de nicel olarak Scene Flow, KITTI 2012 ve KITTI2015 veri kümelerindeki görüntüler kullanılarak değerlendirilmiştir. Aşağıdaki takip eden başlıklarda ilk olarak ağın değişken parametrelerini en iyilemek için yapılan ablasyon çalışmaları verilmektedir (Bölüm 6.2.1). Daha sonra Scene Flow veri seti üzerindeki nitel ve nicel karşılaştırmalar verilmektedir (Bölüm 6.2.2). Son olarak ise KITTI veri kümeleri üzerinde nitel ve nicel karşılaştırmalar verilmektedir (Bölüm 6.2.3).

6.2.1 Ablasyon çalışmaları

Önerilen P3SNet ağının başarımının artırılabilmesi amacıyla ağ üzerinde bazı ablasyon çalışmaları gerçekleştirilmiştir. Bunlardan ilkinde P3 katmanında farklı ölçek değerlerinde öznetelikler elde ederek yapının başarımı test edilmiştir. İkinci ablasyon çalışmasında ise ölçekleme işlemi farklı bir yöntem ile gerçekleştirilerek ağın başarımı üzerindeki etkileri gözlemlenmiştir. Bir diğer ablasyon çalışmasında ise ağın girişinde bulunan öznetelik çıkarma katmanında kullanılan evrişimsel sinir ağlarının çekirdek boyutlarının başarım üzerindeki etkisi incelenmiştir.

İlk ablasyon çalışmasında ölçek değerinin P3SNet ağının başarımı ve hızı üzerindeki etkisini gözlemllemek için bir dizi deneme yapılmıştır. Bu amaçla Şekil 6.1'de verilen P3SNet mimarisinde bulunan P3'ten elde edilen farklı sayıda çıktıya göre farklı ağ yapıları tasarlanmıştır.

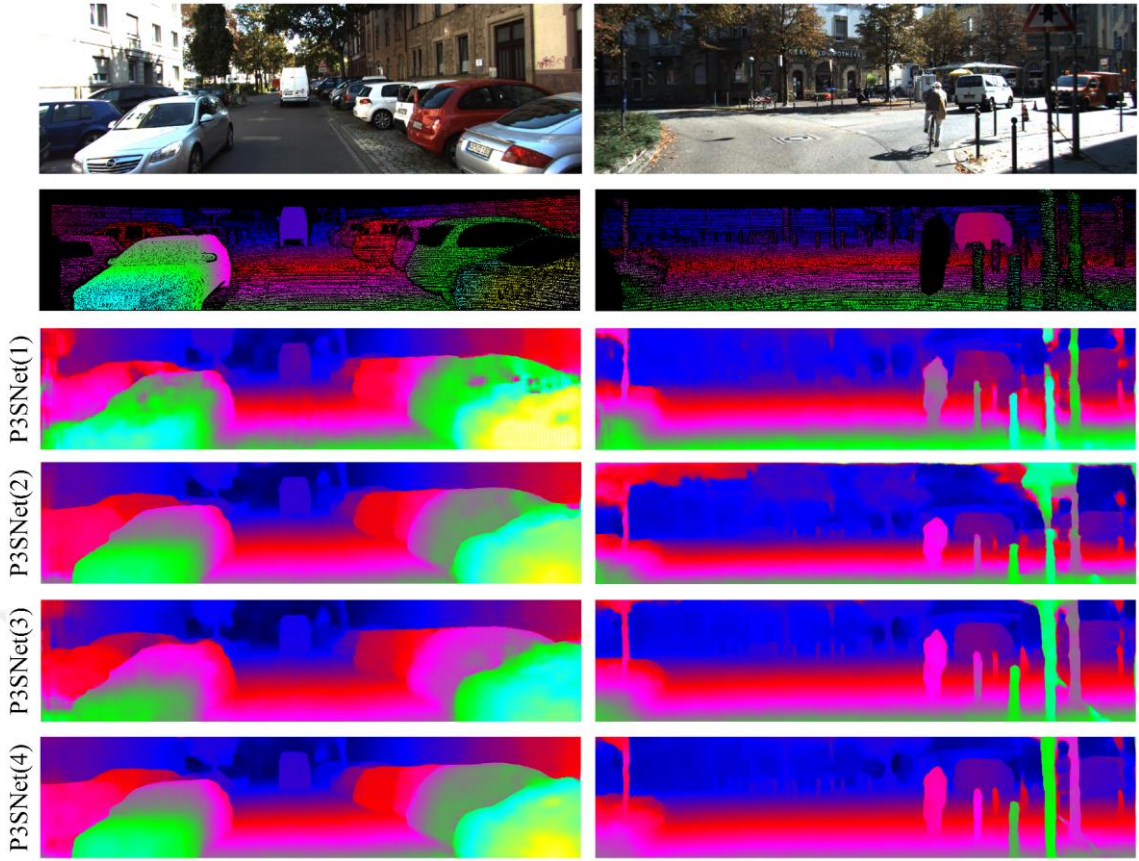
Önermiş olduğumuz P3SNet yapısında farklı ölçek faktörleri için ($n = 1, 2, 3, 4$) bir ablasyon çalışması yapılmıştır. Ablasyon çalışması ile elde edilen sonuçlar Çizelge 6.5'de verilmiştir.

Çizelge 6.5 P3SNet yapısındaki P3 alt katmanının farklı ölçek değerleri için ablasyon çalışması sonuçları

Yöntem	Scene Flow EPE	KITTI		Çalışma süresi (ms)
		KITTI 2012 (>3px)	KITTI 2015 (>3px)	
P3SNet(1)	1.58	5.79	4.85	6.48
P3SNet(2)	1.35	5.05	4.29	9.38
P3SNet(3)	1.16	3,47	3,81	11.73
P3SNet(4)	1.09	3.19	3.73	15.46

Çizelge 6.5’de, "Yöntem" sütununda bulunan parantez içindeki sayılar, ağırlık ölçekleme faktörünü göstermektedir. Çizelge 6.5’de görüldüğü gibi, ölçekleme katmanlarının sayısını artırmak, hem KITTI hem de Scene Flow veri kümelerindeki hata değerlerinin düşmesini sağlamıştır. Aynı zamanda, artan ölçek sayısı ile birlikte beklendiği gibi çalışma süresi de artmaktadır.

P3 alt katmanının farklı ölçek değerlerinin etkisi nicel karşılaştırmalara ek olarak niteliksel olarak da karşılaştırılmıştır. Şekil 6.3'te KITTI 2015 veri kümesinden alınan görüntüler ile P3SNet yapısının farklı ölçek seviyeleri için elde edilen nihai eşitsizlik haritaları gösterilmektedir.



Şekil 6.3. Farklı ölçek değerlerine sahip ağlar tarafından üretilen eşitsizlik haritaları

Şekil 6.3'te de görüldüğü gibi ölçeğin boyutu arttıkça, eşitsizlik haritaları daha doğru ve bağlantılı hale gelmektedir. Ancak sahnedeki nesnelerin kabaca yorumlanması isteniyorsa, daha düşük ölçek seviyelerinde kabul edilebilir doğruluk oranları ile hızlı sonuçlar alındığı görülmektedir. Düşük ölçek değerlerinde elde edilen eşitsizlik haritalarının da nesnelerin genel olarak doğru eşitsizlik değerlerine sahip olduğu görülmektedir.

İkinci ablasyon çalışmasında ise HDA katmanında eşitsizlik haritalarının elde edilmesi işleminde uygulanan ters evrişim işleminin yerine normal evrişim işlemi uygulanarak eşitsizlik haritası elde edildikten sonra üst ölçeğe çıkarma işleminde çift doğrusal aradeğerleme (bilinear interpolation, BI) işlemi kullanılmıştır. Oluşturulan model P3SNet_{bi} olarak adlandırılmıştır. Bir önceki ablasyon çalışmasında olduğu gibi dört farklı ölçekte elde edilen üç piksel hataları ve çalışma süreleri karşılaştırılmıştır, elde edilen sonuçlar Çizelge 6.6'da verilmiştir.

Çizelge 6.6 Çift doğrusal aradeğerleme kullanılarak elde edilen farklı ölçekteki ağların ablasyon çalışması sonuçları

Yöntem	Scene Flow EPE	KITTI		Çalışma süresi (ms)
		KITTI 2012 (>3px)	KITTI 2015 (>3px)	
P3SNet _{bi} (1)	1.83	6.66	5.51	6.29
P3SNet _{bi} (2)	1.62	5.10	4.36	9.12
P3SNet _{bi} (3)	1.32	3.91	4.05	11.37
P3SNet _{bi} (4)	1.20	3.43	3.51	14.94

Çizelge 6.6’da Çizelge 6.5’de olduğu gibi, ölçekleme katmanlarının sayısını artırmak, hem KITTI hem de Scene Flow veri kümelerindeki hata değerlerinin azalmasını sağlamıştır. Aynı zamanda, artan ölçek sayısı ile birlikte beklendiği gibi çalışma sürelerinin arttığı da gözlemlenmiştir. Çizelge 6.6 ve Çizelge 6.5’de elde edilen sonuçlar karşılaştırıldığında P3SNet yapıları ile her bir ölçekte elde edilen sonuçların P3SNet_{bi} yapıları ile elde edilen sonuçlardan daha iyi olduğu görülmektedir. Bu durum, eğitilebilir olmayan bir ölçekleme yöntemi yerine evrişim katmanı kullanmanın ağı eğitimi üzerinde olumlu bir etkiye sahip olduğunu göstermektedir. Fakat P3SNet yapılarında her bir katman eksta bir adet fazla evrişim yapısı içerdiğinden dolayı çalışma süresini artırdığı gözlemlenmiştir.

Üçüncü ablasyon çalışmasında ise sadece öznetelik çıkarma katmanındaki evrişim işlemlerinde çekirdek yapıları değiştirilmiştir. Elde edilen sonuçlar Çizelge 6.7’de verilmiştir. P3SNet_{bi}(3) yapısı karşılaştırma işlemleri için referans olarak alınmıştır ve P3SNet_{bi}(3) yapısında kullanılan işlemler Çizelge 6.7’de Yapı 1 olarak verilmiştir.

Çizelge 6.7’de verilen deneysel çalışmalarda ilk olarak f_1, f_3 ve f_5 isimli katmanlardaki çekirdek boyutları değiştirilerek ideal çekirdek boyutunu belirlemek amaçlanmıştır (Yapı 1, Yapı 2, Yapı 3 ve Yapı 4). Yapı 1 ve Yapı 2 de büyük çekirdek boyutundan başlayıp azalarak devam eden yapıların sabit boyutlu yapılara (Yapı 4) ve küçük çekirdek boyutundan büyük çekirdek boyutuna doğru ilerleyen yapılara (Yapı 3) göre daha iyi bir başarımla ortaya koyduğu gözlemlenmiştir. Yapı 1 ve Yapı 2 başarımla aynı seviyede olduğundan dolayı daha az parametre içeren Yapı 1 üzerinden diğer deneysel çalışmalar gerçekleştirilmiştir.

Çizelge 6.7 Öznitelik çıkarma katmanında evrişimsel sinir ağlarının çekirdek boyutları üzerinde yapılan ablasyon çalışması sonuçları

		Yapı 1	Yapı 2	Yapı 3
Katman İsmi	Katman Girişi	Çekirdek	Çekirdek	Çekirdek
f_1	I_{sol}, I_{sag}	$7 \times 7,2 + BN + ReLU$	$9 \times 9,2 + BN + ReLU$	$3 \times 3,2 + BN + ReLU$
f_2	f_1	$3 \times 3,1 + BN + ReLU$	$3 \times 3,1 + BN + ReLU$	$3 \times 3,1 + BN + ReLU$
f_3	f_2	$5 \times 5,2 + BN + ReLU$	$7 \times 7,2 + BN + ReLU$	$5 \times 5,2 + BN + ReLU$
f_4	f_3	$3 \times 3,1 + BN + ReLU$	$3 \times 3,1 + BN + ReLU$	$3 \times 3,1 + BN + ReLU$
f_5	f_4	$3 \times 3,1 + BN + ReLU$	$5 \times 5,1 + BN + ReLU$	$7 \times 7,1 + BN + ReLU$
f_6	f_5	$3 \times 3,1 + BN + ReLU$	$3 \times 3,1 + BN + ReLU$	$3 \times 3,1 + BN + ReLU$
f_7	f_6	$3 \times 3,1 + BN + ReLU$	$3 \times 3,1 + BN + ReLU$	$3 \times 3,1 + BN + ReLU$
f_8	f_7	$3 \times 3,1 + BN + ReLU$	$3 \times 3,1 + BN + ReLU$	$3 \times 3,1 + BN + ReLU$
EPE		1.32	1.32	1.43

		Yapı 4	Yapı 5	Yapı 6
Katman İsmi	Katman Girişi	Çekirdek	Çekirdek	Çekirdek
f_1	I_{sol}, I_{sag}	$3 \times 3,2 + BN + ReLU$	$7 \times 7,2 + BN + ReLU$	$7 \times 11,2 + BN + ReLU$
f_2	f_1	$3 \times 3,1 + BN + ReLU$	—	$3 \times 3,1 + BN + ReLU$
f_3	f_2	$3 \times 3,2 + BN + ReLU$	$5 \times 5,2 + BN + ReLU$	$5 \times 9,2 + BN + ReLU$
f_4	f_3	$3 \times 3,1 + BN + ReLU$	—	$3 \times 3,1 + BN + ReLU$
f_5	f_4	$3 \times 3,1 + BN + ReLU$	$3 \times 3,1 + BN + ReLU$	$3 \times 5,1 + BN + ReLU$
f_6	f_5	$3 \times 3,1 + BN + ReLU$	—	$3 \times 3,1 + BN + ReLU$
f_7	f_6	$3 \times 3,1 + BN + ReLU$	$3 \times 3,1 + BN + ReLU$	$3 \times 3,1 + BN + ReLU$
f_8	f_7	$3 \times 3,1 + BN + ReLU$	$3 \times 3,1 + BN + ReLU$	$3 \times 3,1 + BN + ReLU$
EPE		1.63	1.46	1.52

Yapı 1 de bulunan f_2 , f_4 ve f_6 isimli katmanların kaldırılması ile Yapı 5 elde edilmiştir. Yapı 1 ile Yapı 5 karşılaştırıldığı zaman hata oranında ciddi artma olduğu gözlemlenmiştir. Bu durum f_2 , f_4 ve f_6 isimli katmanların öznitelik çıkarma işlemindeki olumlu etkisini göstermektedir. Yapı 6 ise f_1 , f_3 ve f_5 katmanlarında uygulanan evrişim işleminin yatay eksenindeki arama düzlemi dikey eksene göre uzatılarak elde edilmiştir. Çizelge 6.7’de verilen deneysel sonuçlar doğrultusunda bu çalışma nezninde yatay çekirdeklerdeki boyutlarda oluşturulan artışın başarımlar üzerinde olumlu bir etkisi olmadığı gözlemlenmiştir.

Bu sonuçlar göz önünde bulundurarak, P3SNet(3) ve P3SNet(4) yapıları tezin ilerleyen bölümlerindeki deneysel çalışmalarda kullanmak için seçilmiştir. Tezin ilerleyen kısımlarında bu yapılar P3SNet (n=3) ve P3SNet+ (n=4) olarak adlandırılmıştır.

P3SNet ve P3SNet+ ağ yapılarının literatürdeki diğer yöntemlerle niteliksel ve niceliksel olarak karşılaştırmalarının yanı sıra ayrıca çalışma süreleride ele alınarak karşılaştırmalar ve değerlendirmeler yapılmıştır. Önerilen ağ yapıları ile ilgili olarak karşılaştırma işlemleri ilk olarak Scene Flow veri kümesi üzerinde gerçekleştirilmiştir.

6.2.2 Scene Flow veri kümesinde sonuçların karşılaştırma

Önerilen P3SNet ağ yapısında nihai eşitsizlik haritası elde edilirken ağın ara katmanlarında da farklı ölçeklerde eşitsizlik haritaları üretilmektedir. P3SNet ağ yapısının ve yapıdaki katmanların öğrenme yeteneklerini gözlemlemek için her bir katmandan elde edilen eşitsizlik haritaları Scene Flow test veri kümesinde incelenmiştir. Bu karşılaştırmalarda EPE hatası kullanılarak toplam ortalama hata oranları Çizelge 6.8’de verilmiştir. Ayrıca Çizelge 6.8’de P3SNet ağının ara ölçeklerdeki eşitsizlik haritalarının elde edilme süreleri de verilmektedir.

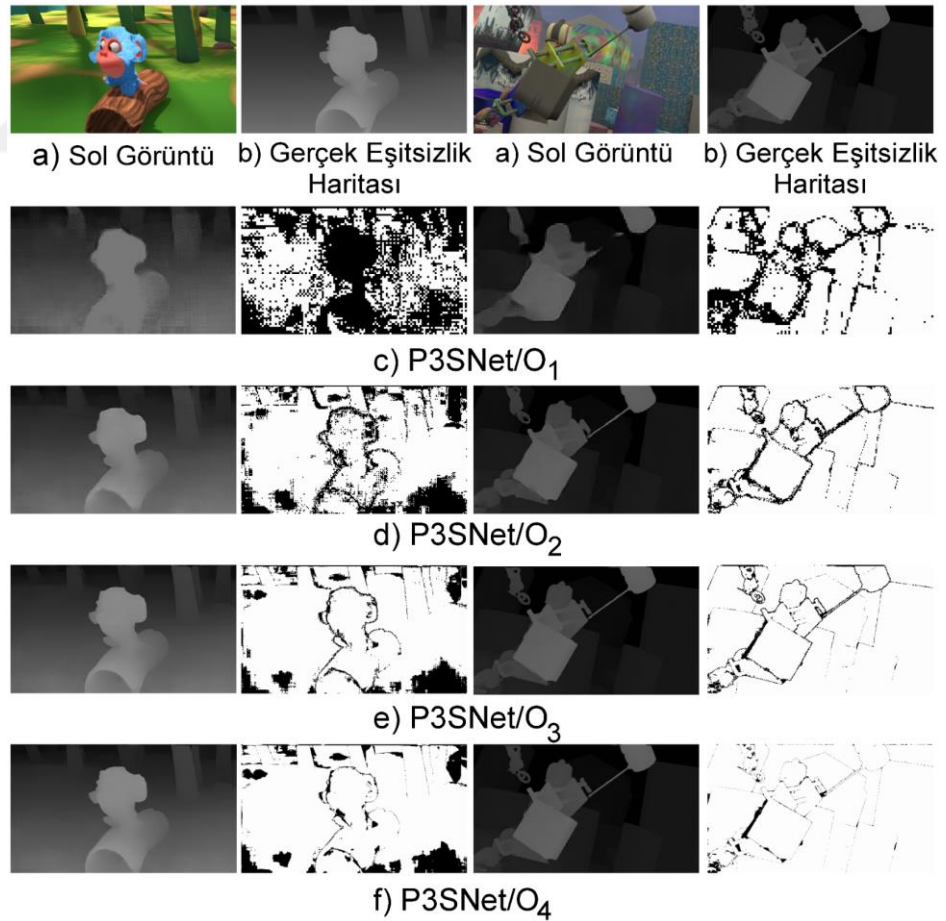
Çizelge 6.8’de verilen O_4 , ağ yapısının giriş görüntüsü ile aynı boyutlara sahip olan nihai eşitsizlik haritasına ait çıkışı temsil etmektedir. Tablodaki gösterimlerde yer alan O_i ifadesi ağın i ’ninci ($i = 1, 2, 3$ ve 4) çıkış görüntüsünü temsil etmektedir. Bu çıkışların elde edildiği yerler görsel olarak Şekil 6.1’de gösterilmektedir. Örneğin O_3 ifadesinden elde edilen çıkış boyutu orijinal boyutun hem derinlik hem de yükseklik olarak yarısına eşit olan eşitsizlik haritasını temsil etmektedir.

Çizelge 6.8. Farklı ölçeklerdeki eşitsizlik haritalarının ortalama EPE hataları ve elde edilme süreleri

Çıkış Düğümü	EPE	Süre (ms)
O_1	4.62	7.74
O_2	2.14	9.15
O_3	1.40	10.91
O_4	1.16	11.73

Çizelge 6.8’de verilen EPE sonuçlarına göre, her bir katmandan elde edilen hata oranlarının önemli oranda azaldığı görülmektedir. Bu durum, elde edilen eşitsizlik haritalarının doğruluğunun her bir ölçekte iyileştiğini göstermektedir. Aynı zamanda bu işlemler esnasındaki çalışma süreleri arasındaki farklar karşılaştırıldığında, katmanlar arasındaki işlem sürelerinin çok fazla olmadığı görülmektedir.

Şekil 6.4’de Scene Flow veri kümesindeki görüntülerin önerilen P3SNet ağ yapısına giriş olarak verilmesi ile farklı katmanlardan elde edilen eşitsizlik haritalarının karşılaştırmaları verilmektedir. Şekil 6.4’de ilk satır birinci ve üçüncü sütundaki görüntüler orijinal sol giriş görüntülerini, ikinci ve dördüncü sütundaki görüntüler ise doğru eşitsizlik haritalarını temsil etmektedir. Diğer satırlarda yer alan görüntülerde sol sütunda bulunan görüntüler, yukarıdan aşağıya doğru ağın ilk katmanından son katmanına doğru elde edilen eşitsizlik haritalarını göstermektedir. Sağ sütunda yer alan görsellerde ise ilgili ölçekte elde edilen eşitsizlik haritalarına ait üç piksel hata haritaları gösterilmektedir.



Şekil 6.4. Farklı boyutlarda elde edilen eşitsizlik ve hata haritaları

Şekil 6.4c’de görüldüğü gibi P3SNet ağının ilk ürettiği eşitsizlik haritası oldukça bulanık bir görüntü yapısına sahipken, hata oranının da yüksek olduğu görülmektedir. Fakat bu görüntüde nesnelerin ana eşitsizlik değerlerinin genellikle doğru olduğu, genel olarak kenar ve geçiş bölgelerindeki eşitsizlik değerlerinde hatalar olduğu gözlenmektedir. Şekil 6.4d ile bir önceki ölçekteki Şekil 6.4c’de yer alan eşitsizlik haritası kıyaslandığında Şekil 6.4d’de bulunan eşitsizlik haritasının biraz daha netleştiği görülmektedir. Ayrıca Şekil 6.4d’de bulunan hata haritasındaki hatalar Şekil 6.4c’de bulunan hata haritasına göre önemli ölçüde azalmaktadır. Buradaki hata haritası üzerindeki hataların artık çoğunun görüntülerin sınır bölgelerinde yer aldığı gözlemlenmektedir (Şekil 6.3d). Benzer şekilde, Şekil 6.4e ve Şekil 6.4f’de görüldüğü üzere eşitsizlik haritasının boyutları büyütüldüğü zaman görüntülerdeki netliklerin arttığı görülmektedir. Şekil 6.3’de bulunan hata haritaları karşılaştırıldığında her bir ölçek artışında sınır bölgelerindeki hataların azaldığı gözlemlenmektedir. Bu durum ise ağın eğitiminde temel olarak alınan kabadan inceye bir eğitim yapısının başarılı bir şekilde gerçekleştirildiğini göstermektedir.

P3SNet ağ yapısının literatürdeki önde gelen diğer yöntemler ile karşılaştırılması Scene Flow veri kümesindeki test veri kümesinde gerçekleştirilmiştir. Elde edilen sonuçlar Çizelge 6.9’da verilmektedir.

Çizelge 6.9. Scene Flow test setinde performans karşılaştırma sonuçları

	MC-CNN	GANet	GC-Net	SegStereo	CRL	PSMNet	LEAStereo	*DP-Best
EPE	3.79	0.84	2.51	1.45	1.32	1.09	0.78	0.86
Süre(s)	67	1.8	0.90	0.6	0.47	0.41	0.3	0.182
	AAANet	*DP-Fast	DispNetC	MSCVNet	StereoNet	P3SNet+	P3SNet	
EPE	0.87	0.97	1.68	1.32	1.1	1.09	1.16	
Süre(s)	0.07	0.062	0.06	0.041	0.015	0.015	0.012	

*DP, DeepPruner'in kısaltmasıdır.

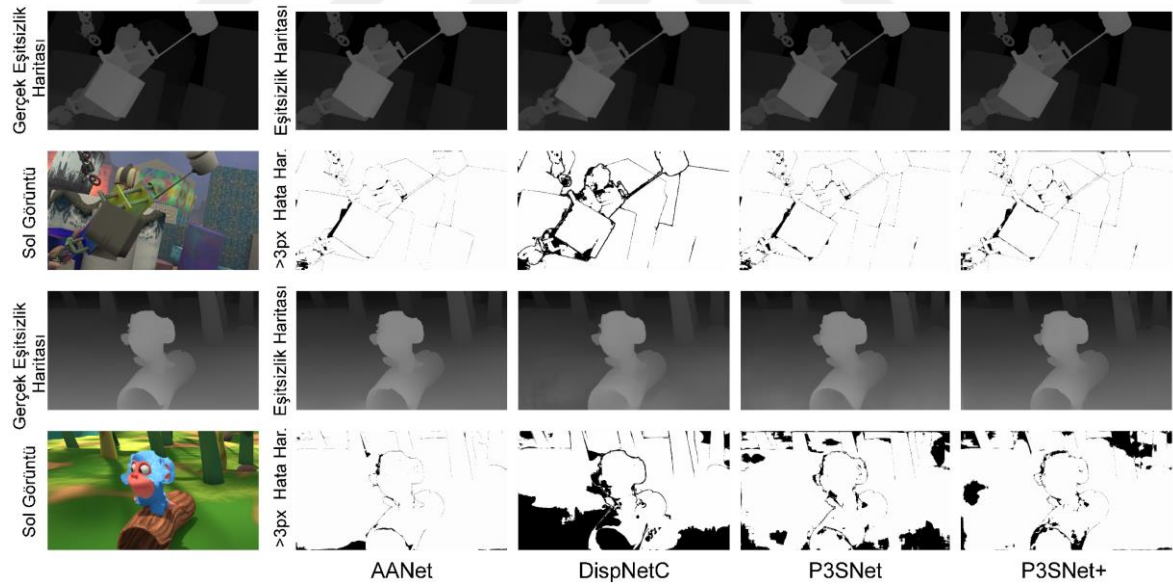
Çizelge 6.9’da üst grupta gerçek zamanlı olmayan önde gelen yöntemler verilirken alttaki ikinci grupta gerçek zamanlı çalışmalar için uygun olabilecek yöntemler verilmektedir. Çizelge 6.9’da görüldüğü gibi P3SNet algoritması kendisinden neredeyse beş kat daha yavaş çalışma süresine sahip olan DispNetC yönteminden ve neredeyse 4 kat daha yavaş çalışma süresine sahip olan MSCVNet (Jia vd., 2021) daha düşük hata değerleri üretmektedir. P3SNet algoritması StereoNet (Khamis vd., 2018) ile karşılaştırıldığında ise çalışma süresi olarak %20 daha iyi durumda iken hata oranında %5 daha kötü durumdadır. StereoNet yapısı

P3SNet+ ile karşılaştırıldığında çalışma süresi olarak birbirlerine oldukça yakın iken önerilen yöntemin başarısının biraz daha iyi olduğu görülmektedir.

Aynı zamanda P3SNet ve P3SNet+ algoritmaları, literatürde yüksek başarımlı olarak bilinen ve ağır işlem yüklerine sahip SegStereo, MC-CNN CRL ve GC-Net gibi yöntemleri de başarılarıyla geride bırakmıştır.

GANet, DP, PSMNet ve AANet yöntemleri ile kıyaslandığı zaman ise hata değerindeki fark ile çalışma süreleri göz önüne alındığında P3SNet ve P3SNet +ağ yapılarının daha verimli olduğu açık bir şekilde görülmektedir.

Scene Flow veri kümesi üzerindeki nitel karşılaştırmalar için DispNetC ve AANet yöntemleri seçilmiştir. Şekil 6.5’de hem eğitim kümesinden hem de test kümesinden bir görüntü ile gerçekleştirilen görsel karşılaştırmalar verilmiştir. Bu görüntülerden ilki test kümesinden ikincisi ise eğitim kümesinden seçilmiştir.



Şekil 6.5. Scene Flow veri kümesi üzerindeki nitel karşılaştırmalar

Şekil 6.5’de verilen karşılaştırmalar da ilk sütundaki görüntüler ağlara giriş olarak verilen stereo görüntülerden soldaki kameraya ait görüntüleri ve doğru eşitsizlik haritalarını göstermektedir. İkinci sütundan beşinci sütuna kadar olan görüntüler ise sırayla AANet DispNetC, P3SNet ve P3SNet+ yöntemlerinden elde edilen eşitsizlik haritalarını ve bu eşitsizlik haritalarına ait üç piksel hata haritalarını göstermektedir.

Şekil 6.5'deki görsellerde hem eşitsizlik haritaları hem de hata haritaları karşılaştırıldığı zaman P3SNet ve P3SNet+ ağ yapıları ile üretilen sonuçların DispNetC ile elde edilen eşitsizlik haritalarından çok daha başarılı olduğu görülmektedir. İlk test görüntüsündeki hata haritaları karşılaştırıldığı zaman nesnelerin sınır bölgelerindeki hataların belirgin bir oranda azaldığı gözlemlenmektedir. İkinci test görüntüsündeki karşılaştırmalarda ise DispNetC ağının hata haritasının sol bölgesinde bulunan büyük bir alanda yanlış eşleşmeler yaptığı gözlenirken, P3SNet ve P3SNet+ ağ yapılarının bu bölgelerde daha doğru sonuçlar verdiği gözlemlenmektedir. Özellikle P3SNet+ ağının bu bölgedeki sonuçlarının oldukça doğru olduğu görülmektedir.

AANet ile elde edilen sonuçlar karşılaştırıldığı zaman ise ilk test görüntüsünde başarımların birbirine yakın olduğu görülmektedir. İkinci test görüntüsünde ise P3SNet+ ile edilen sonucun P3SNet ile elde edilen sonuçtan iyi olduğu görülmektedir. Fakat AANet ile elde edilen eşitsizlik haritasının önerdiğimiz her iki yöntemle elde edilen eşitsizlik haritalarından daha başarılı olduğu görülmektedir. Öte yandan çalışma süreleri karşılaştırıldığında önerdiğimiz P3SNet ve P3SNet+ ağ yapılarının AANet'den sırayla yaklaşık 6 ve 5 kat daha hızlı olduğu göz önünde bulundurulduğunda, bu hata oranlarının kabul edilebilir seviyede olduğu görülmektedir.

6.2.3 KITTI veri kümelerinde sonuçların karşılaştırması

KITTI 2012 ve KITTI 2015 veri kümelerinde de Scene Flow veri kümesinde olduğu gibi hem nitel hem de nicel karşılaştırmalar yapılmıştır. Nicel karşılaştırmalar için yüzdesel kötü eşleşen piksellerin hata oranı metriği kullanılmıştır. Bu hata metriği kullanılırken KITTI 2012 veri kümesinde 2, 3, 4 ve 5 olmak üzere dört farklı hata eşik değeri kullanılmıştır. KITTI 2015 veri kümesinde ise sadece 3 piksellik eşik değeri kullanılarak hata oranları hesaplanmıştır.

Çizelge 6.10'da önerilen P3SNet yapısı ile literatürdeki önde gelen stereo eşleşme modellerinin KITTI 2012 çevrimiçi test seti üzerindeki karşılaştırılması verilmektedir.

Çizelge 6.10. KITTI 2012 çevrimiçi test setinden elde edilen karşılaştırma sonuçları

Yöntem	>2 piksel		>3 piksel		>4 piksel		>5 piksel		Süre (s)
	noc	all	noc	all	noc	all	noc	all	
MC-CNN	3.9	5.45	2.43	3.63	1.9	2.85	1.64	2.39	67
GANet	1.89	2.5	1.19	1.6	0.91	1.23	0.76	1.02	1.8
GC-Net	2.71	3.46	1.77	2.3	1.36	1.77	1.12	1.46	0.9
Content-CNN	4.98	6.51	3.07	4.29	2.39	3.36	2.03	2.82	0.7
SegStereo	2.66	3.19	1.68	2.03	1.25	1.52	1.00	1.21	0.6
PSMNet	2.44	3.01	1.49	1.89	1.12	1.42	0.9	1.15	0.41
LEAStereo	1.90	2.39	1.13	1.45	0.83	1.08	0.67	0.88	0.3
AAnet	2.90	3.60	1.91	2.42	1.46	1.87	1.20	1.53	0.06
DispNetC	7.38	8.11	4.11	4.65	2.77	3.20	2.05	2.39	0.06
Flow2Stereo	6.56	7.32	4.58	5.11	3.61	4.03	3.00	3.35	0.05
StereoNet	-	-	4.91	6.02	-	-	-	-	0.015
P3SNet	6.44	7.46	3.65	4.46	2.51	3.18	1.91	2.48	0.012
P3SNet+	6.27	7.37	3.55	4.42	2.43	3.12	1.85	2.41	0.015

Çizelge 6.10’da görüldüğü üzere P3SNet ağ yapısının, çalışma süresi olarak StereoNet’den sonra kendisine en yakın olan Flow2Stereo (Liu vd., 2020) yönteminden dört kattan daha fazla hızlı olduğu görülmektedir. Elde edilen hata oranlarına bakıldığında ise P3SNet ve P3SNet+ ağ yapılarının, Flow2Stereo ağından bütün hata değerlerinde daha iyi olduğu görülmektedir. P3SNet ve P3SNet+ ağ yapıları DispNetC ile karşılaştırıldığında ise 5 piksel (all) kıyaslaması hariç bütün değerlerde iyi olduğu görülmektedir. StereoNet’in bilinen değerleri için önerilen ağlarla karşılaştırıldığında önerilen ağların daha başarılı olduğu görülmektedir.

İşlem yükü açısından oldukça ağır olan yöntemlerle hata oranları karşılaştırıldığında ise P3SNet yönteminin hata oranlarının yüksek olduğu görülmektedir. Fakat çalışma süresi göz önüne alındığında, P3SNet ağı Çizelge 6.10’da verilen yüksek başarılı ağlardan en hızlı olanı (LEAStereo) ile karşılaştırıldığında zaman 25 kat, en yavaş olanı (MC-CNN) ile karşılaştırıldığında zaman ise 3722 kat daha hızlı olduğu görülmektedir.

Çizelge 6.11’de ise önerilen P3SNet yapısının KITTI 2015 çevrimiçi test seti üzerindeki karşılaştırmaları verilmektedir.

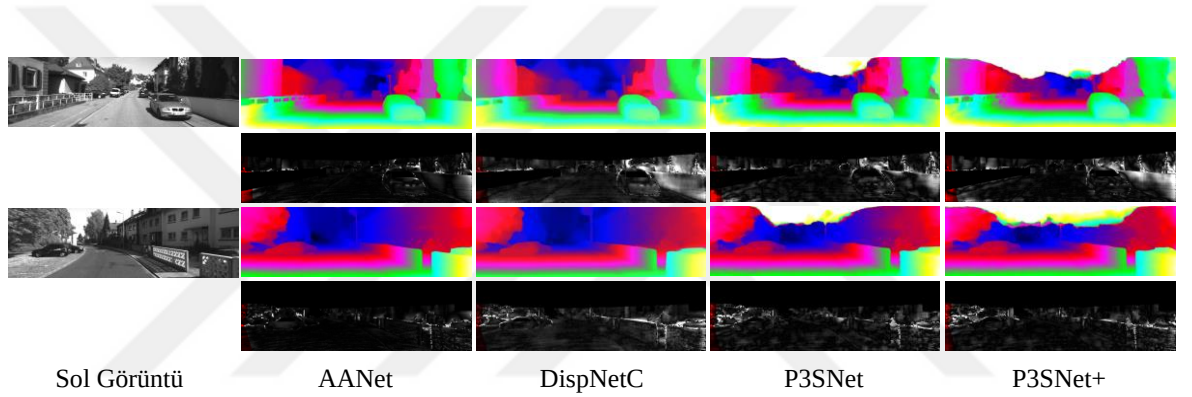
Çizelge 6.11. KITTI 2015 çevrimiçi test setinden elde edilen karşılaştırma sonuçları

Yöntem	all			noc			Süre (s)
	D1-bg	D1-fg	D1-all	D1-bg	D1-fg	D1-all	
MC-CNN	2.89	8.88	3.89	2.48	7.64	3.33	67
GANet	1.48	3.46	1.81	1.34	3.11	1.63	1.8
GC-Net	2.21	6.16	2.87	2.02	5.58	2.61	0.9
Content-CNN	3.73	8.58	4.54	3.32	7.44	4.00	0.7
SegStereo	1.88	4.07	2.25	1.76	3.70	2.08	0.6
CRL	2.48	3.59	2.67	2.32	3.12	2.45	0.47
PSMNet	1.86	4.62	2.32	1.71	4.31	2.14	0.41
LEAStereo	1.40	2.91	1.65	1.29	2.65	1.51	0.3
PCStereo	2.39	4.98	2.82	2.23	4.65	2.63	0.2
AAnet	1.99	5.39	2.55	1.80	4.93	2.32	0.06
DispNetC	4.32	4.41	4.34	4.11	3.72	4.05	0.06
Flow2Stereo	5.01	14.62	6.61	4.77	14.03	6.29	0.05
DeepCostAggr	5.34	11.35	6.34	4.82	10.11	5.69	0.034
SSBSR	-	-	-	4.83	10.75	5.67	0.034
StereoNet	4.30	7.45	4.83	-	-	-	0.015
P3SNet	4.4	8.28	5.05	4.1	7.44	4.65	0.012
P3SNet+	4.15	7.59	4.72	3.86	6.66	4.32	0.015

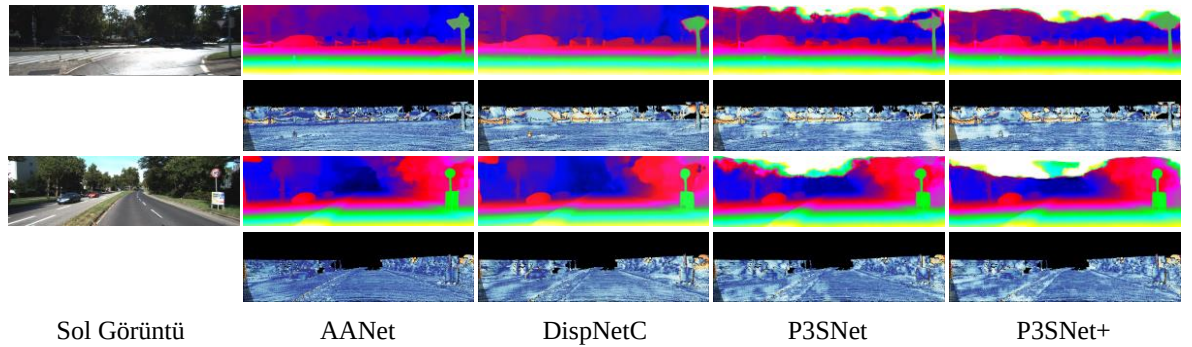
Çizelge 6.11’de verilen sonuçlarda P3SNet yapısının, StereoNet’den sonra kendisine en yakın çalışma süresine sahip olan DeepCostAggr (Kuzmin vd., 2017) ve SSBSR (Miclea ve Nedeveschi, 2020) yöntemlerine göre iki katdan daha fazla hızlı olduğu görülmektedir. Ayrıca P3SNet ve P3SNet+’ın bütün karşılaştırma parametrelerinde Flow2Stereo, DeepCostAggr ve SSBSR yapılarından daha düşük hata oranları elde ettiği görülmektedir. Özellikle ön plan sahneleri için verilen hata oranları incelendiğinde, P3SNet ve P3SNet+ yapılarının DeepCostAggr, Flow2Stereo ve SSBSR yapılarına göre dikkate değer farklar ile daha iyi olduğu görülmektedir. P3SNet ve P3SNet+ yapılarının DispNetC ağı ile karşılaştırıldığı zaman ise arka plan sahnelerindeki hata oranları birbirine yakın olmasına karşın ön plan sahne alanlarında DispNetC ağı daha başarılıdır. P3SNet+ yapısı yaklaşık aynı çalışma süresine sahip StereoNet’in bilinen değerleri ile karşılaştırıldığında D1-fg (all) hariç daha başarılı olduğu görülmektedir. P3SNet ise StereoNet’den %20 daha az çalışma süresine sahip olmasına karşın hata değerleri oldukça yakındır.

Çizelge 6.11’de, Çizelge 6.10’da olduğu gibi PCStero (Chuah vd., 2022) gibi gerçek zamanlı uygulamalar için çalışma süresi uzun yöntemlerin hata oranları karşılaştırıldığı zaman ise P3SNet ve P3SNet+ yöntemlerinin hata oranlarının yüksek olduğu görülmektedir.

Nitel karşılaştırmalar için Scene Flow veri kümesinde olduğu gibi AANet ve DispNetC yöntemleri seçilmiştir. Şekil 6.6 ve Şekil 6.7’de KITTI 2012 ve KITTI 2015 çevrimiçi test setinden alınan iki görüntü kullanılarak görsel karşılaştırmalar verilmiştir. Şekil 6.6 ve Şekil 6.7’de sol sütundaki ilk görüntüler stereo kameradan soldaki kameraya ait olan giriş görüntüleridir. Diğer sütunlardaki görüntüler sırayla AANet, DispNetC, P3SNet ve P3SNet+ ağlarından elde edilen eşitsizlik haritalarına aittir. Bu görüntülerin altında da bu haritalara ait hata haritaları verilmektedir.



Şekil 6.6. KITTI 2012 çevrimiçi test setindeki nitel karşılaştırmalar



Şekil 6.7. KITTI 2015 çevrimiçi test setindeki nitel karşılaştırmalar

Şekil 6.6 ve Şekil 6.7’de verilen görseller incelendiğinde AANet ile elde edilen eşitsizlik haritalarının en doğru sonuçları verdiği görülmektedir. Fakat çalışma süresi göz önüne alındığında P3SNet yönteminin AANet’den yaklaşık 6 kat daha hızlı olduğu ve P3SNet+’ında AANet’den yaklaşık 5 kat daha hızlı olduğu göz önünde bulundurulmalıdır.

DispNetC ile her iki yöntemle elde edilen eşitsizlik haritaları karıştırıldığında biri diğerinden üstün demek mümkün değildir. DispNetC ağından elde edilen eşitsizlik haritalarındaki bazı yanlış eşleşmiş bölgeler önerdiğimiz P3SNet ve P3SNet+ ağ yapılarında doğru şekilde eşleştirilebilmişken tam tersi durumlar da söz konusudur.

Özellikle ağların çalışma süreleri ve elde ettikleri hata oranları karşılaştırıldığı zaman, önerilen P3SNet ve P3SNet+ ağ yapılarının oldukça başarılı sonuçlar verdiği gözlemlenmektedir.

Sonuç olarak, bu çalışmada eşitsizlik haritası elde etmek amacıyla uçtan uca bir ağ yapısı (P3SNet) önerilmiştir. Önerilen P3SNet ve P3SNet+ ağ yapıları tasarlanırken üretilen eşitsizlik haritalarının gerçek zamanlı uygulamalarda kullanılabilir olması göz önünde bulundurulmuştur. Nihayetinde yalnızca 12ms ve 15ms’de eşitsizlik haritası üretebilen iki adet ağ yapısı sunulmuştur.

Bu ağ yapıları tasarlanırken öznitelik çıkarma aşaması için paralel piramit havuzu (P3) yapısı önerilmiştir. Bu yapı tasarlanırken geleneksel öznitelik çıkarma yöntemi olan SIFT’in ölçek uzayı oluşturma yapısından esinlenilmiştir. P3 yapısı ile farklı ölçeklerde eşleşme maliyetleri üretilmiştir. Elde edilen farklı ölçeklerdeki eşleşme maliyetleri ile kabadan inceye öğrenme yapısı olan bir ağ tasarlanmıştır. P3 yapısı ile küresel ve bölgesel özniteliklerden daha iyi faydalanılarak eşitsizlik haritası üretilmesi amaçlanmıştır.

P3SNet ağının ana eğitimi Scene Flow veri kümesinde gerçekleştirilmiştir KITTİ 2012 ve KITTİ 2015 veri kümelerinde ise ince ayar işlemleri gerçekleştirilmiştir.

Elde edilen sonuçlarda önerilen P3SNet ve P3SNet+ ağ yapıları ile gerçek zamanlı uygulamalar için kabul edilebilir bir başarıya sahip eşitsizlik haritaları elde edilebilmektedir.

BÖLÜM VII

SONUÇ

Bu tez kapsamında ilk olarak stereo eşleştirme yöntemleri ile eşitsizlik haritalarının elde edilmesi ve düzeltilmesine yönelik yapılan çalışmalarla ilgili ayrıntılı literatür çalışması sunulmuştur. Literatürde gerçekleştirilen çalışmalar incelenerek katkı sağlanması amacıyla, stereo görüntüler kullanılarak eşitsizlik haritalarının düzeltilmesi ve eşitsizlik haritalarının elde edilmesine yönelik farklı ağ yapıları geliştirilmiştir. Tez kapsamında gerçekleştirilen çalışmalardan eşitsizlik haritalarının ve eşleşme maliyetlerinin düzeltilmesi amacıyla;

- Emlek ve Peker (2020)’de yayınlanan eşitsizlik haritalarının düzeltilmesi için SimRef ağ yapısı (Bölüm 3)
- Emlek ve Peker (2021a)’da piksellerin arama düzlemi boyunca içerdikleri sıralı bilgilerinin eşleşme maliyetlerinin düzeltilmesinde kullanılabilmesi için CR-RNN ağ yapısı (Bölüm 4)

önerilmiştir.

Uçtan uca ağ yapıları ile eşitsizlik haritalarının elde edilmesine yönelik;

- Emlek ve Peker (2021b)’de stereo eşleştirme işlemlerinde öznitelik çıkarma katmanının başarısını artırmaya yönelik yatay dikkatli eşitsizlik ağı (Bölüm 5)
- Geleneksel öznitelik çıkarma yöntemleri olan SIFT algoritmasından esinlenerek tasarlanan öznitelik çıkarmanı kullanılarak P3SNet yapısı (Bölüm 6)

önerilmiştir.

Yukarıda ifade edilen çalışmalar ayrıntılı olarak incelendiğinden ilk olarak Bölüm 3’te eşleşme maliyetlerini elde etmenin mümkün olmadığı durumlar için eşitsizlik haritalarının iyileştirilmesini amaçlayan SimRef ağ yapısı önerilmiştir. SimRef ağıyla sadece sol kamera görüntüsü ve Content-CNN (Luo vd., 2016) ağı ile elde edilen eşitsizlik haritaları kullanarak düzeltme işlemi gerçekleştirmiştir. Elde edilen sonuçlarda, önerilen ağın çok ağır bir işlem yükü oluşturmadan gerçek zamanlı sistemlere entegre edilebilecek

şekilde eşitsizlik haritalarının iyileştirildiği gösterilmiştir. Ayrıca SimRef yöntemi, mevcut stereo görüntü tabanlı eşitsizlik bilgisini kullanan sistemler içerisinde performansı arttırmak için doğrudan kullanılabilir yapıya sahiptir. Böylece, stereo görüntülerin kullanıldığı ve eşitsizlik bilgisine ihtiyaç duyulan uygulamalarda da başarımların artırılmasına olanak sağlanacaktır.

Bölüm 4’de piksellerin ve dolayısıyla eşleşme maliyetlerinin sıralı yapısı kullanılarak eşleşme maliyetlerinin düzeltilmesi amacıyla CR-RNN yapısı önerilmiştir. Bu ağ yapısı giriş olarak aldığı eşleşme maliyetlerindeki sıralı bilgileri RNN ağı ile anlamlı hale getirilmiştir. Yapılan çalışmada literatürdeki oldukça popüler olarak kullanılan üç farklı metottan elde edilen eşleşme maliyetleri kullanılarak ağın başarısı gösterilmiştir. Kullanılan bu metotlardan derin öğrenme tabanlı eşleşme maliyeti üretme yöntemi olan MC-CNN-Hızlı yöntemine CR-RNN ağı ile görece birkaç parametre eklenerek ağın başarımı MC-CNN-Doğru yöntemine yaklaştırılmıştır. Ayrıca, CR-RNN’nin başarısı, CBCA ve SAD gibi geleneksel yöntemlerle elde edilen eşleşme maliyetleri kullanılarak da gösterilmiştir. Nicel karşılaştırmalarda CR-RNN’nin KITTI 2012 ve KITTI 2015 veri kümelerinde üç farklı metotla elde edilen eşleşme maliyetleri üzerinde doğruluğu artırdığı gösterilmiştir. Benzer şekilde, niteliksel sonuçlar eşitsizlik haritalarındaki iyileştirmeleri daha görünür kılmaktadır. Bu sonuçlar, önerilen yöntemin nesne sınırlarında ve aynı zamanda yollar gibi düşük dokulu alanlarda eşitsizlik değerlerini iyileştirmek için kullanılabilirliğini kanıtlamaktadır. Önerilen yöntem basit bir yapıya sahiptir ve bu nedenle, çıktı olarak eşleşme maliyetleri veren yöntemlere kolayca uyarlanabilir.

Bölüm 5’te stereo eşleştirme işleminde piksellerin sıralı yapısını kullanmak amacıyla yatay dikkatli evrişim (HAC) katmanı önerilmiştir. Önerilen HAC yapısının literatürde kabul görmüş olan DispNetC ağının öznetelik çıkarma katmanına uyarlanması ile yatay dikkatli eşitsizlik ağı (HADN) elde edilmiştir. HAC yapısının DispNetC ağı üzerindeki etkisini gözlemlemek için Scene Flow veri kümesi üzerinde nitel ve nicel karşılaştırmalar yapılmıştır. Nicel karşılaştırmalar sonunda HAC yapısının DispNetC ağının başarısını yaklaşık %13 oranında artırdığı gözlemlenmiştir. HAC yapısı ile DispNetC ağı nitel olarak karşılaştırıldığında ise, HAC yapısının eşitsizlik haritaları üzerinde önemli derecede doğrulukları arttığı gözlemlenmiştir. Özellikle HAC yapısı ile eşitsizlik haritalarındaki nesnelerin kenar bölgelerinde daha keskin ve doğru sonuçlar elde edilmiştir. Ayrıca bu çalışmada eklenen harici iki evrişim işleminin ise çalışma süresi

üzerinde önemli derecede bir etkisi olmamıştır. HAC yapısı derin öğrenme tabanlı bütün stereo eşleştirme yöntemlerindeki öznitelik çıkarma işlemlerinde basitçe kullanılabilecek bir yapıdır. Bu yapının DispNetC ağı üzerindeki etkisi bu çalışma ile gösterilmiştir.

Bölüm 6’da, eşitsizlik haritası elde etmek amacıyla uçtan uca bir ağ yapısı (P3SNet ve P3SNet+) önerilmiştir. Önerilen P3SNet ve P3SNet+ ağ yapıları tasarlanırken tezin hedeflerinden biri olan gerçek zamanlı uygulamalarda kullanılabilirlik göz önünde tutularak üretilen eşitsizlik haritalarının elde edilme süreleri mümkün olduğunca küçük tutulmaya çalışılmıştır. Bu ağ yapıları tasarlanırken öznitelik çıkarma aşaması için paralel piramit havuzu (P3) yapısı önerilmiştir. P3 yapısı tasarlanırken geleneksel öznitelik çıkarma yöntemi olan SIFT’in ölçek uzayı oluşturma yapısından esinlenilmiştir. P3 yapısı ile farklı ölçeklerde eşleşme maliyetleri üretilmiştir. Elde edilen farklı ölçeklerdeki eşleşme maliyetleri ile kabadan inceye öğrenme yapısı olan bir ağ tasarlanmıştır. P3 yapısı ile küresel ve bölgesel özniteliklerden daha iyi faydalanılarak eşitsizlik haritası üretilmesi amaçlanmıştır. P3SNet ağının ana eğitimi Scene Flow veri kümesinde gerçekleştirilmiştir. KITTI 2012 ve KITTI 2015 veri kümelerinde ise ince ayar işlemleri gerçekleştirilmiştir. Elde edilen sonuçlarda önerilen P3SNet ve P3SNet+ ağ yapıları ile gerçek zamanlı uygulamalar için kabul edilebilir bir başarıya sahip eşitsizlik haritaları elde edilebilmektedir.

Tezin hedefleri doğrultusunda tez çalışması kapsamında gerçek zamanlı uygulamalar için eşitsizlik haritalarının elde edilmesi ve düzeltilmesine yönelik çalışmalar gerçekleştirilmiştir. Eşitsizlik haritalarının ve eşleşme maliyetlerinin düzeltilmesi amacıyla önerilen yöntemlerde yatay düzlemlerdeki piksel yönelim bilgileri kullanılarak başarımlar artırılmıştır. P3SNet ağ yapısı ile küresel ve bölgesel içerik bilgilerinin daha iyi çıkarılmasını sağlayan bir modül içeren uçtan uca bir ağ yapısı tasarlanmıştır. Bu ağ yapısı ile hem gerçek zamanlı hem de kabul edilebilir bir başarıya sahip eşitsizlik haritalarının elde edilmesi sağlanmıştır.

KAYNAKLAR

Adhyapak, S., Kehtarnavaz, N. and Nadin, M., “Stereo matching via selective multiple windows”, *Journal of Electronic Imaging*, 16(1), 013012:1-14, 2007.

Ambrosch, K., Kubinger, W., Humenberger, M. and Steininger, A. “Hardware implementation of an SAD based stereo vision algorithm”, *2007 IEEE Conference on Computer Vision and Pattern Recognition*, Minneapolis, MN, USA, s. 1-6, 17-22 June, 2007.

Baker, S., Scharstein, D., Lewis, J. P., Roth, S., Black, M. J. and Szeliski, R., “A database and evaluation methodology for optical flow”, *International Journal of Computer Vision*, 92(1), 1-31, 2011.

Banks, J., Bennamoun, M., and Corke, P., “Non-parametric techniques for fast and robust stereo matching”, *IEEE Region 10 International Conference TENCON*, Brisbane, Australia, s. 365-368, 4 December, 1997.

Bleyer, M. and Gelautz, M., “Graph-cut-based stereo matching using image segmentation with symmetrical treatment of occlusions”, *Signal Processing: Image Communication*, 22(2), 127-143, 2007.

Brandao, P., Mazomenos, E. and Stoyanov, D., “Widening siamese architectures for stereo matching”, *Pattern Recognition Letters*, 120, 75-81, 2019.

Buades, A. and Facciolo, G., “Reliable multiscale and multiwindow stereo matching”, *SIAM Journal on Imaging Sciences*, 8(2), 888-915, 2015.

Chang, J. R. and Chen, Y. S., “Pyramid stereo matching network”, *2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, s. 5410-5418, 18-23 June, 2018.

Chen, J., Benzeroual, K. and Allison, R. S., “Calibration for high-definition camera rigs with marker chessboard”, *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Providence, RI, USA, s. 29-36, 16-21 June, 2012.

Chen S., Xiang Z., Qiao C., Chen Y., and Bai T., "PGNet: Panoptic parsing guided deep stereo matching", *Neurocomputing*, 463(C), 609–622, 2021.

Chen, Z., Sun, X., Wang, L., Yu, Y. and Huang, C., “A deep visual correspondence embedding model for stereo matching costs”, *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, s. 972-980, 7-13 December, 2015.

Cheng X., Zhong, Y., Harandi, M., Dai, Y., Chang, X., Li, H., Drummond, T. and Ge, Z., "Hierarchical Neural Architecture Search for Deep Stereo Matching", *Advances in Neural Information Processing Systems*, 33, 1-12, 2020.

Chuah, W., Tennakoon, R., Hoseinnezhad, R. and Bab-Hadiashar, A., "Deep Learning-Based Incorporation of Planar Constraints for Robust Stereo Depth Estimation in Autonomous Vehicle Applications", *IEEE Transactions on Intelligent Transportation Systems*, 23(7), 6654-6665, 2022.

Dinh, V. Q., Pham, C. C. and Jeon, J. W., “Robust adaptive normalized cross-correlation for stereo matching cost computation”, *IEEE Transactions on Circuits and Systems for Video Technology*, 27(7), 1421-1434, 2017.

Duggal S., Wang S., Ma W. C., Hu R. and Urtasun R., "DeepPruner: Learning Efficient Stereo Matching via Differentiable PatchMatch", *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea, s. 4383-4392, 27 October.-2 November, 2019.

Dosovitskiy, A., Fischer, P., Ilg, E., Hausser, P., Hazirbas, C., Golkov, P., Smagt, V. D., Cremers, D. and Brox, T., “FlowNet: learning optical flow with convolutional networks”, *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, s. 2758-2766, 7-13 December, 2015.

Dong, C., Loy, C. C., He, K. and Tang, X., “Image super-resolution using deep convolutional networks”, *IEEE Transactions On Pattern Analysis And Machine Intelligence*, 38(2), 295-307, 2015.

Dong, C., Loy, C. C., He, K. and Tang, X., “Learning a deep convolutional network for image super-resolution”, *European Conference On Computer Vision*, Zurich, Switzerland, s. 184-199, 6-12 September 2014.

Dong, Q. and Feng, J., “Outlier detection and disparity refinement in stereo matching”, *Journal of Visual Communication and Image Representation*, 60, 380-390, 2019.

Emlek, A. ve Peker, M., “Evrişimli sinir ağı tabanlı derinlik haritası iyileştirme”, *Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi*, 9(2), 722-729, 2020.

Emlek, A. and Peker, M., Refinement of matching costs for stereo disparities using recurrent neural networks”, *J Image Video Proc.*, 11, 1-19, 2021a.

Emlek, A. ve Peker, M., “Stereo Eşleştirme için Yatay Dikkatli Evrişim Katmanı”. **29. Sinyal İşleme ve İletişim Uygulamaları Kurultayı**, İstanbul, Türkiye, s. 1-4, 9-11 Haziran, 2021b.

Ernst, I. and Hirschmüller, H., “Mutual information based semi-global stereo matching on the GPU”, *4th International Symposium on Visual Computing*, Las Vegas, NV, USA, s. 228-239, 1-3 December, 2008.

Fookes, C., Bennamoun, M. and Lamanna, A., “Improved stereo image matching using mutual information and hierarchical prior probabilities”, *2002 International Conference on Pattern Recognition*, Quebec City, Canada, s. 937-940, 11-15 August, 2002.

Forstmann, S., Kanou, Y., Ohya, J., Thuerling, S. and Schmitt, A., “Real-time stereo by using dynamic programming”, *Conference on Computer Vision and Pattern Recognition Workshop*, Washington, DC, USA, s. 29-29. 27 June-2 July, 2004.

Geiger, A., Lenz, P. and Urtasun, R., “Are we ready for autonomous driving? The KITTI vision benchmark suite”, *2012 IEEE Conference on Computer Vision and Pattern Recognition*, Providence, RI, USA, s. 3354-3361, 16-21 June, 2012.

Geiger, A., Roser, M. and Urtasun, R. “Efficient large-scale stereo matching”, *Asian Conference on Computer Vision*, Queenstown, New Zealand, s. 25-38, 8-12 November, 2010.

Girshick, R., Donahue, J., Darrell, T. and Malik, J., “Rich feature hierarchies for accurate object detection and semantic segmentation”, *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, s. 580-587, 23-28 June, 2014.

Gong, M., Yang, R., Wang, L., and Gong, M., “A performance study on different cost aggregation approaches used in real-time stereo matching”, *International Journal of Computer Vision*, 75(2), 283-296, 2007.

Guan, S., Klette, R. and Woo, Y. W., “Belief propagation for stereo analysis of night-vision sequences”, *Pacific-Rim Symposium on Image and Video Technology (PSIVT 2009)*, Heidelberg, Berlin, s. 932-943, 13-16 January, 2009.

Guo, B., Dai, H., Li, Z. and Huang, W., "Efficient planar surface-based 3d mapping method for mobile robots using stereo vision", *IEEE Access*, 7, 73593-73601, 2019.

Guo, X., Yang, K., Yang, W., Wang, X. and Li, H., “Group-Wise correlation stereo network”, *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, s. 3268-3277, 15-20 June, 2019.

Hamzah, R. A., Abd Rahim, R. and Noh, Z. M. “Sum of absolute differences algorithm in stereo correspondence problem for stereo matching in computer vision application” *3rd International Conference on Computer Science and Information Technology*, Chengdu, China, s. 652-657, 9-11 July, 2010.

Hayashi, K., Kato, H. and Nishida, S., “Occlusion detection of real objects using contour based stereo matching”, *2005 International Conference on Augmented Tele-Existence*,

Christchurch, New Zealand, s. 180-186, 5-8 December, 2005.

He, K., Sun, J. and Tang, X., “Guided image filtering”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(6), 1397-1409, 2013.

He, K., Sun, J. and Tang, X., “Guided image filtering”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(6), 1397-1409, 2012.

Heo, Y. S., Lee, K. M. and Lee, S. U., “Robust stereo matching using adaptive normalized cross-correlation”, *IEEE Transactions on pattern analysis and machine intelligence*, 33(4), 807-822, 2010.

Hirschmuller, H., “Accurate and efficient stereo processing by semi-global matching and mutual information”, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, San Diego, CA, USA, s. 807-814, 20-25 June, 2005.

Hirschmuller, H. and Scharstein, D., “Evaluation of cost functions for stereo matching”, *IEEE Conference on Computer Vision and Pattern Recognition*, Minneapolis, MN, USA, s. 1-8, 17-22 June, 2007.

Hosni, A., Bleyer, M., Gelautz, M. and Rhemann, C., “Local stereo matching using geodesic support weights”, *16th IEEE International Conference on Image Processing (ICIP)*, Cairo, Egypt, s. 2093- 2096, 7-10 November, 2009.

Hosni A., Bleyer, M. , Rhemann, C., Gelautz M. and Rother, C., “Real-time local stereo matching using guided image filtering”, *2011 IEEE International Conference on Multimedia and Expo*, Barcelona, Spain, s. 1-6, 11-15 July, 2011.

Jia, X., Chen, W., Li, C., Liang, Z., Wu, M., Tan, Y. and Huang, L., “Multi-Scale Cost Volumes Cascade Network for Stereo Matching”, *2021 IEEE International Conference on Robotics and Automation (ICRA)*, Xi'an, China, s. 8657-8663, 30 May - 05 June, 2021.

Jiao, J., Wang, R., Wang, W., Dong, S., Wang, Z. and Gao, W., “Local stereo matching with improved matching cost and disparity refinement”, *IEEE MultiMedia*, 21(4), 16-27, 2014.

Kanade, T. and Okutomi, M., “A stereo matching algorithm with an adaptive window: theory and experiment”, *1991 IEEE International Conference on Robotics and Automation*, Sacramento, CA, USA, s. 1088-1095, 9-11 April, 1991.

Kang, J., Chen, L., Deng, F. and Heipke, C., “Context pyramidal network for stereo matching regularized by disparity gradients”, *ISPRS Journal of Photogrammetry and Remote Sensing*, 157, 201-215, 2019.

Kendall, A., Martirosyan, H., Dasgupta, S., Henry, P., Kennedy, R., Bachrach, A. and Bry, A., “End-to-end learning of geometry and context for deep stereo regression”, *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy, s. 66-75 2017, 22-29 October, 2017.

Khamis, S., Fanello, S., Rhemann, C., Kowdle, A., Valentin, J., Izadi, S., “StereoNet: Guided Hierarchical Refinement for Real-Time Edge-Aware Depth Prediction”, *European Conference on Computer Vision 2018 (ECCV 2018)*, Munich, Germany, s. 596–613, 8-14 September, 2018.

Kim, J. C., Lee, K. M., Choi, B. T. and Lee, S. U., “A dense stereo matching using two-pass dynamic programming with generalized ground control points”, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, CA, USA, s. 1075-1082, 20-25 June, 2005.

Kim, K. R., Koh, Y. J. and Kim, C. S., “Multiscale feature extractors for stereo matching cost computation”, *IEEE Access*, 6, 27971-27983, 2018.

Kim, S., Min, D., Kim, S. and Sohn, K., “Unified Confidence Estimation Networks for Robust Stereo Matching”, *IEEE Transactions on Image Processing*, 28(3), 1299-1313, 2019.

Kingma, D. P. and Ba, J. L., “Adam: A method for stochastic optimization”, ***3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings***, San Diego, USA, s. 1-13, 7 - 9 May, 2015.

Kiureghian A. D. and Ditlevsen O., “Aleatory or epistemic? Does it matter?”, ***Structural Safety***, 31(2), 105-112, 2009.

Koh, Y. J. and Kim, C. S., “Primary object segmentation in videos based on region augmentation and reduction”, ***2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)***, Honolulu, HI, USA, s. 7417-7425, 21-26 July 2017.

Kolmogorov, V. and Zabih, R., “Graph cut algorithms for binocular stereo with occlusions”, ***Handbook of Mathematical Models in Computer Vision***, Boston, USA, s. 423-437, 2006.

Kuzmin, A., Mikushin, D., and Lempitsky, V., "End-to-End learning of cost-volume aggregation for real-time dense stereo", ***2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP)***, Tokyo, Japan, s. 1-6, 25-28 September, 2017.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken A., Tejani, A., Totz, J., Wang, Z. and Shi, W., “Photo-realistic single image super-resolution using a generative adversarial network”, ***IEEE conference on computer vision and pattern recognition***, Honolulu, HI, USA, s. 4681-4690, 21-26 July, 2017.

Liang, Z., Feng, Y., Guo, Y., Liu, H., Chen, W., Qiao, L., Zhou, L. and Zhang, J., “Learning for disparity estimation through feature constancy”, ***2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition***, Salt Lake City, UT, USA , s. 2811-2820, 18-23 June, 2018.

Lim ,S., Mittal, A., Davis, L. S. and Paragios, N., “Uncalibrated stereo rectification for automatic 3D surveillance”, ***2004 International Conference on Image Processing***, Singapore, s. 1357-1360, 24-27 June, 2004.

Liu, P., King, I., Lyu, M. R. and Xu, J., "Flow2Stereo: Effective Self-Supervised Learning of Optical Flow and Stereo Matching", **2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)**, Seattle, WA, USA, s. 6647-6656, 13-19 June, 2020.

Long, J., Shelhamer, E. and Darrell, T., "Fully convolutional networks for semantic segmentation", **2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**, Boston, MA, USA, s. 3431-3440, 7-12 June, 2015.

Lowe, D.G. "Distinctive image features from scale-invariant keypoints", **International Journal of Computer Vision**, 60(2), 91-110, 2004.

Lu, B., Sun, L., Yu, L. and Dong, X., "An improved graph cut algorithm in stereo matching", **Displays**, 69(102052), 1-7, 2021.

Luo, W., Schwing, A .G. and Urtasun, R., "Efficient deep learning for stereo matching", **2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**, Las Vegas, NV, USA, s. 5695-5703, 27-30 June, 2016.

Ma, Z., He, K., Wei, Y., Sun, J. and Wu, E., "Constant time weighted median filtering for stereo matching and beyond", **2013 IEEE International Conference on Computer Vision**, Sydney, NSW, Australia, s. 49-56, 1-8 December, 2013.

Marshall, M. R., Hellfeld, D., Joshi, T. H. Y., Salathe, M., Bandstra, M. S., Bilton, K. J., Cooper, R. J., Curtis, J. C., Negut, V., Shurley, A. J. and Vetter, K., "3-D object tracking in panoramic video and LiDAR for radiological source-object attribution and improved source detection", **IEEE Transactions on Nuclear Science**, 68(2), 189-202, 2021.

Mattoccia, S., Tombari, F. and Di Stefano, L., "Stereo Vision Enabling Precise Border Localization Within a Scanline Optimization Framework", **Asian Conference on Computer Vision ACCV 2007**, Berlin, Heidelberg, s. 517-527, 18-22 November, 2007.

Mayer, N., Ilg, E., Hausser, P., Fischer, P., Cremers, D., Dosovitskiy, A. and Brox, T., "A large dataset to train convolutional networks for disparity, optical flow, and scene flow

estimation”, *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, s. 4040-4048, 27-30 June, 2016.

Menze, M. and Geiger, A., “Object scene flow for autonomous vehicles”, *2015 IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, s. 3061-3070, 7-12 June, 2015.

Michailidis, G., Pajarola, R. and Andreadis, I., “High Performance Stereo System for Dense 3-D Reconstruction”, *IEEE Transactions on Circuits and Systems for Video Technology*, 24(6), 929-941, 2014.

Miclea, V. C. and Nedevschi, S., "Real-Time Semantic Segmentation-Based Stereo Reconstruction", *IEEE Transactions on Intelligent Transportation Systems*, 21(4), 1514-1524, 2020.

Ohta, Y. and Kanade, T., “Stereo by intra- and inter-scanline search using dynamic programming”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(2), 139-154, 1985.

Okutomi, M. and Kanade, T., “A multiple-baseline stereo”, *IEEE Transactions on pattern analysis and machine intelligence*, 15(4), 353-363, 1993.

Otte, M. and Nagel, H. H., “Optical flow estimation: Advances and comparisons”, *European Conference on Computer Vision*, Berlin, Heidelberg, s. 49-60, 2 - 6 May, 1994.

Pang, J., Sun, W., Ren, J.SJ., Yang, C. and Yan, Q., “Cascade residual learning: a two-stage convolutional neural network for stereo matching”, *IEEE International Conference on Computer Vision Workshops, ICCVW 2017*, Venice, Italy, s. 878-886, 22-29 October, 2017.

Peng, Y., Li, G., Wang, R. and Wang, W., “Stereo matching with space-constrained cost aggregation and segmentation-based disparity refinement”, *Three-Dimensional Image Processing, Measurement (3DIPM), and Applications*, 939309, 1-11, 2015.

Perri, S., Colonna, D., Zicari, P. and Corsonello, P. "SAD-based stereo matching circuit for FPGAs", *13th IEEE International Conference on Electronics, Circuits and Systems*, Nice, France, s. 846-849, 10-13 December 2006.

Ren, S., He, K., Girshick, R. and Sun, J., "Faster R-CNN: Towards real-time object detection with region proposal networks", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), 1137-1149, 2017.

Rhemann, C., Hosni, A., Bleyer, M., Rother, C. and Gelautz, M., "Fast cost-volume filtering for visual correspondence and beyond". *Conference on Computer Vision and Pattern Recognition (CVPR) 2011*, Colorado Springs, CO, USA s. 3017-3024, 20-25 June, 2011.

Scharstein, D., Hirschmüller, H., Kitajima, Y., Krathwohl, G., Nešić, N., Wang, X. and Westling, P., "High-resolution stereo datasets with subpixel-accurate ground truth", *Pattern Recognition*, 31-42, 2014.

Scharstein, D. and Szeliski, R., "A Taxonomy and evaluation of dense two-frame stereo correspondence algorithms", *International Journal of Computer Vision*, 47(1), 7-42, 2002.

Seki, A. and Pollefeys, M., "SGM-Nets: Semi-global matching with neural networks", *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, s. 6640-6649, 21-26 July, 2017.

Shen, S., "Accurate Multiple View 3D Reconstruction Using Patch-Based Stereo for Large-Scale Scenes", *IEEE Transactions on Image Processing*, 22(5), 1901-1914, 2013.

Shi, F., Hughes, N.R. and Roberts, G., "SSD Matching Using Shift-Invariant Wavelet Transform", *12th British Machine Vision Conference*, Manchester, UK, s. 1-10, 10-13 September, 2001.

Shi, H., Zhu, H., Wang, J., Yu, S. Y. and Fu, Z.F., "Segment-based adaptive window and multi-feature fusion for stereo matching" *Journal of Algorithms & Computational*

Technology, 10(1), 3-11, 2016.

Song, X., Zhao, X., Hu, H. and Fang, L., “EdgeStereo: A Context Integrated Residual Pyramid Network for Stereo Matching” *Asian Conference on Computer Vision ACCV 2018*, Perth, Australia, s. 20–35, 2–6 December, 2018.

Sun, J., Zheng, N. N. and Shum, H. Y., “Stereo matching using belief propagation”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(7), 787-800, 2003.

Sun, P., Zhao, X., Xu, Z., Wang R. and Min, H., "A 3D LiDAR data-based dedicated road boundary detection algorithm for autonomous vehicles", *IEEE Access*, 7, 29623-29638, 2019.

Sun, Y., Wang, X. and Tang, X., “Deep learning face representation from predicting 10,000 classes”, *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, Ohio, s. 1891-1898, 24-27 June, 2014.

Taigman, Y., Yang, M., Ranzato, M. and Wolf, L., “DeepFace: Closing the gap to human-level performance in face verification”, *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, s. 1701-1708, 23-28 June, 2014.

Tang, Z. and Hwang, J., "MOANA: An online learned adaptive appearance model for robust multiple object tracking in 3D", *IEEE Access*, 7, 31934-31945, 2019.

Wang C., Wang X., Zhang J., Zhang L., Bai X., Ning X., Zhou J. and Edwin H., “Uncertainty estimation for stereo matching based on evidential deep learning”, *Pattern Recognition*, 124, 108498:1-13, 2022.

Wang, L., Gong, M., Gong, M. and Yang, R., “How far can we go with local optimization in real-time stereo matching”, *Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06)*, Chapel Hill, USA, s. 129-136, 14-16 June, 2006.

Weon, I. S., Lee, S. G. and Ryu, J. K., "Object recognition based interpolation with 3D LIDAR and vision for autonomous driving of an intelligent vehicle", *IEEE Access*, 8, 65599-65608, 2020.

Williem, and Park, I. K. "Deep self-guided cost aggregation for stereo matching", *Pattern Recognition Letters*, 112, 168-175, 2018.

Xie, S. and Tu, Z. "Holistically-Nested edge detection". *International Journal of Computer Vision*, 125(1-3), 3-18, 2017.

Xu H. and Zhang J., "AANet: Adaptive Aggregation Network for Efficient Stereo Matching", *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, s. 1956-1965, 13-19 June, 2020.

Yan, T., Gan, Y., Xia, Z. and Zhao, Q., "Segment-based disparity refinement with occlusion handling for stereo matching", *IEEE Transactions on Image Processing*, 28(8), 3885-3897, 2019.

Yang, G., Song, X., Huang, C., Deng, Z., Shi, J. and Zhou, B., "DrivingStereo: A large-scale dataset for stereo matching in autonomous driving scenarios", *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, s. 899-908, 15-20 June, 2019.

Yang, G., Zhao, H., Shi, J., Deng Z. and Jia, J., "SegStereo: Exploiting Semantic Information for Disparity Estimation", *European Conference on Computer Vision (ECCV 2018)*, Munich, Germany, s. 660–676, 8-14 September 2018.

Yang, Q., Wang, L., Yang, R., Wang, S., Liao, M. and Nister, D., "Real-time Global Stereo Matching Using Hierarchical Belief Propagation", *British Machine Vision Conference 2006*, Edinburgh, UK, s. 989-998, 4-7 September, 2006.

Yoon, K. J. and Kweon, S. I., "Adaptive support-weight approach for correspondence search", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(4), 650-656, 2006.

Žbontar, J. and Lecun, Y., “Computing the stereo matching cost with a convolutional neural network”, **2015 IEEE Conference on Computer Vision and Pattern Recognition**, Boston, MA, USA, s. 1592-1599, 7-12 June, 2015.

Zhang, F., Prisacariu, V., Yang, R., and Torr, P. "GA-Net: Guided aggregation net for end-to-end stereo matching", **2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)**, Long Beach, CA, USA, s. 185-194, 15-20 June, 2019.

Zhang F., Qi X., Yang R., Prisacariu V., Wah B., Torr P. “Domain-Invariant Stereo Matching Networks”, **European Conference on Computer Vision ECCV 2020**, Glasgow, UK, s. 420-439, 23-28 August, 2020.

Zhang, K., Lu, J. and Lafruit, G., “Cross-Based local stereo matching using orthogonal integral images”, **IEEE Transactions on Circuits and Systems for Video Technology**, 19(7), 1073-1079, 2009.

Zhang, K., Lu, J., Lafruit, G., Lauwereins, R. and Van Gool, L. “Robust stereo matching with fast normalized cross-correlation over shape-adaptive regions”, **16th IEEE International Conference on Image Processing**, Cairo, Egypt, s. 2357-2360, 7-10 November, 2009.

Zhang, R., Isola, P. and Efros, A. A., “Colorful image colorization”, **European Conference on Computer Vision (ECCV 2016)**, Amsterdam, Netherlands, s. 649-666, 8-16 October, 2016.

Zhang, R., Zhu, J. Y., Isola, P., Geng, X., Lin, A.S., Yu, T. and Efros, A. A., “Real-Time user-guided image colorization with learned deep priors”, **ACM Transactions on Graphics**, 36(4), 119:1-11, 2017.

Zhang, Y., Chen, Y., Bai, X., Yu, S., Yu, K., Li, Z. and Yang, K., “Adaptive unimodal cost volume filtering for deep stereo matching”, **AAAI Conference on Artificial Intelligence**, New York, USA, s. 12926-12934, 7-12 February, 2020.

Zhan, Y., Gu, Y., Huang, K., Zhang, C. and Hu, K., “Accurate image-guided stereo matching with efficient matching cost and disparity refinement”, *IEEE Transactions on Circuits and Systems for Video Technology*, 26(9), 1632-1645, 2015.

Zhao, X., Sun, P., Xu, Z., Min H. and Yu, H., "Fusion of 3D LIDAR and camera data for object detection in autonomous vehicle applications", *IEEE Sensors Journal*, 20(9), 4901-4913, 2020.

Zhou, X. Z., Wen, G. J. and Wang, R. S., “Fast stereo matching using adaptive window”, *Chinese Journal of Computers*, 29(3), 473-479, 2006.

Zhu, Z. and Dai, Q., “Hybrid scheme for accurate stereo matching”, *Neurocomputing*, 252, 24-33, 2017.

Zinner, C., Humenberger, M., Ambrosch, K. and Kubinger, W., “An optimized software-based implementation of a census-based stereo matching algorithm”, *International Symposium on Visual Computing*, Las Vegas, USA, s. 216-227, 1-3 December, 2008.

ÖZ GEÇMİŞ

Alper EMLEK tarihinde’de doğdu. İlk, orta ve lise öğrenimini Niğde’de tamamladı. 2009 yılında girdiği Niğde Üniversitesi Elektrik-Elektronik Mühendisliği Bölümü’nden Haziran 2014’te mezun oldu. Ağustos 2014’te Niğde Üniversitesi Elektrik-Elektronik Mühendisliği Bölümü’nde yüksek lisans öğrenimine başladı. Ocak 2015’de Niğde Üniversitesi Elektrik-Elektronik Mühendisliği Bölümü’ne araştırma görevlisi olarak atandı ve Temmuz 2017’da yüksek lisans öğrenimini tamamladı. Ağustos 2017’da Niğde Ömer Halisdemir Üniversitesi Elektrik-Elektronik Mühendisliği Bölümü’nde doktora öğrenimine başladı.

TEZ ÇALIŞMASINDAN ÜRETİLEN ESERLER

Bu tez çalışmasından, 1 (bir) adet uluslararası makale, 1 (bir) adet ulusal makale ve 1 (bir) adet ulusal bildiri üretilmiştir. Belirtilen yayınlara ait ayrıntılı bilgi aşağıda sunulmuştur.

Emlek, A. and Peker, M., “Refinement of matching costs for stereo disparities using recurrent neural networks”, *J Image Video Proc.* 11, 1-19, 2021.

Emlek, A. ve Peker, M. “Evrişimli sinir ağı tabanlı derinlik haritası iyileştirme”, *Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi* 9(2), 722-729, 2020.

Emlek, A. ve Peker, M. “Stereo Eşleştirme için Yatay Dikkatli Evrişim Katmanı”. **29. Sinyal İşleme ve İletişim Uygulamaları Kurultayı.** s. 1-4, 9-11 Haziran, 2021

