

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

BİLGİSAYARLA GÖRME VE MAKİNE ÖĞRENMESİ
YÖNTEMLERİ KULLANILARAK MARKET
RAFLARINDA ÜRÜN TANIMA

Ceren Gülra MELEK

DOKTORA TEZİ

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Programı

Danışman

Prof. Dr. Songül VARLI

Eş Danışman

Doç. Dr. Elena BATTİNİ SÖNMEZ

Mart, 2024

T.C.

YILDIZ TEKNİK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

**BİLGİSAYARLA GÖRME VE MAKİNE ÖĞRENMESİ
YÖNTEMLERİ KULLANILARAK MARKET RAFLARINDA
ÜRÜN TANIMA**

Ceren Gülra MELEK tarafından hazırlanan tez çalışması 08.03.2024 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı, Bilgisayar Mühendisliği Programı **DOKTORA TEZİ** olarak kabul edilmiştir.

Prof. Dr. Songül VARLI
Yıldız Teknik Üniversitesi
Danışman

Doç. Dr. Elena BATTİNİ SÖNMEZ
İstanbul Bilgi Üniversitesi
Eş-Danışman

Jüri Üyeleri

Prof. Dr. Songül VARLI, Danışman
Yıldız Teknik Üniversitesi

Prof. Dr. Mehmet Fatih AMASYALI, Üye
Yıldız Teknik Üniversitesi

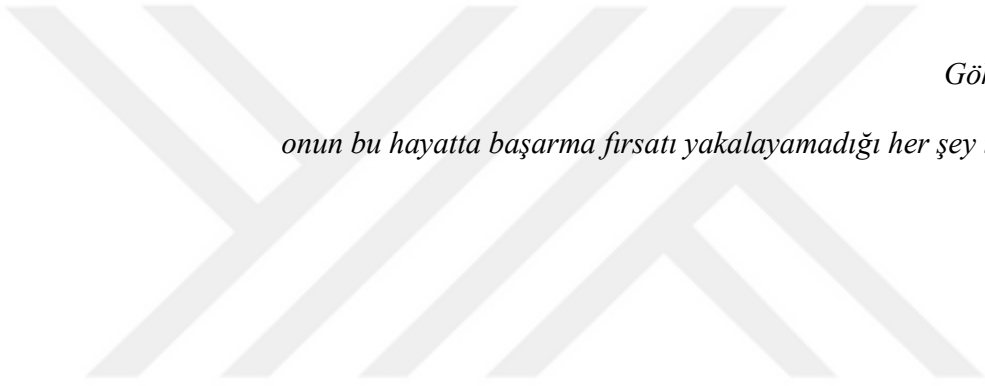
Dr. Öğr. Üyesi Hakan AYRAL, Üye
Yıldız Teknik Üniversitesi

Doç. Dr. Pelin GÖRGEL, Üye
İstanbul Üniversitesi-Cerrahpaşa

Dr. Öğr. Üyesi Oğuz ALTUN, Üye
Yıldız Teknik Üniversitesi

Danışmanım Prof. Dr. Songül VARLI sorumluluğunda tarafımca hazırlanan “Bilgisayarla Görme ve Makine Öğrenmesi Yöntemleri Kullanılarak Market Raflarında Ürün Tanıma” başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

Ceren Gülra MELEK



*Gökay'a,
onun bu hayatta başarma fırsatı yakalayamadığı her şey için...*

TEŞEKKÜR

Tez çalışma sürecimde bilimsel bilgi ve desteğini esirgemeyen danışmanım Prof. Dr. Songül VARLI'ya, tüm çalışma sürecim boyunca kıymetli vaktini ayırıp, fikir ve görüşleriyle bana yol gösteren eş-danışmanım Doç. Dr. Elena BATTİNİ SÖNMEZ'e ve Dr. Öğr. Üyesi Hakan AYRAL'a, kıymetli görüşleri ile çalışmama katkı sağlayan tezimin jüri üyesi Prof. Dr. Mehmet Fatih AMASYALI'ya sonsuz teşekkürlerimi sunarım.

Akademik hayata birlikte başladığım, doktora yolculuğunu birlikte kat ettiğim ve destekleri ile hep yanımda olan değerli arkadaşlarıma; bilgileri, deneyimleri ve duruşlarıyla bana ilham kaynağı olan tüm hocalarıma; manevi destekleri ve bana olan sonsuz güvenleriyle hep yanımda olduklarını hissettiğim canım geniş aileme içtenlikle teşekkür ederim.

Bilgisi, tecrübesi ve çalışkanlığı ile hep örnek olan, bana olan sonsuz güveni ve desteği ile akademisyen olmamdaki en büyük etken babam Dr. Öğr. Üyesi Mustafa MELEK'e; hayatım boyunca desteklerini ve sevgilerini hep hissettiğim annem Gülsen MELEK'e ve abim Edip Candaş MELEK'e; ailenin akademisyenleri olarak aynı yolda yürüdüğümüz ve her zaman birbirimize destek olduğumuz Dr. Öğr. Üyesi Ece MELEK ve Öğr. Gör. Evrim Ceyda MELEK'e; her geçen gün büyümelerine tanıklık etmekten zevk aldığım, sevgi ve enerjileriyle her zaman motivasyon kaynağım olan yeğenlerim Elisa Deniz MELEK'e ve Elina Derin MELEK'e çok teşekkür ederim.

Ceren Gülra MELEK

İÇİNDEKİLER

TEŞEKKÜR	iv
SİMGE LİSTESİ	vii
KISALTMA LİSTESİ	viii
ŞEKİL LİSTESİ	xi
TABLO LİSTESİ	xii
ÖZET	xiv
ABSTRACT	xvi
1 GİRİŞ	1
1.1 Tezin Amacı	1
1.2 Özgün Değer	4
1.3 Literatür Taraması.....	5
1.3.1 Üç Aşamalı Uçtan-uca Olmayan Ürün Tanıma Çalışmaları	8
1.3.2 Ürün Tespit Aşamasını Ele Alan Çalışmalar	9
1.3.3 Ürün Sınıflandırma Aşamasını Ele Alan Çalışmalar	10
1.3.4 Ürün Tespiti ve Ürün Sınıflandırma Aşamalarının Ortak Olarak Gerçekleştiren Çalışmalar	12
1.3.5 Çok Aşamalı Uçtan-uca Ürün Tanıma Çalışmaları	14
1.4 Hipotez	18
1.5 Tezin Genel Bakışı.....	19
2 MARKET RAFLARINDA ÜRÜN TANIMA	21
2.1 Market Raflarında Ürün Tanıma Problemi	21
2.2 Market Raflarında Ürün Tanıma Veri Setleri	25
2.2.1 WebMarket	26
2.2.2 Grozi-120	26
2.2.3 Grocery Products	27
2.2.4 Grocery Dataset	31
2.2.5 Freiburg Groceries	31
2.2.6 CAPG-GP	32

2.2.7	SKU-110K	32
2.2.8	Locount	33
2.2.9	Unitail-Det	33
2.3	Market Raflarında Ürün Tanıma için Kullanılan Bilgisayarla Görme ve Makine Öğrenmesi Yöntemleri.....	34
2.3.1	Ürün Tespiti için Kullanılan Yöntemler	34
2.3.2	Özellik Çıkarma Yöntemleri	37
2.3.3	Özelliklerin Sınıflandırılması için Kullanılan Yöntemler	43
2.3.4	Sonuçların İyileştirilmesi için Kullanılan Yöntemler	45
3	ÇOK AŞAMALI UÇTAN-UCA ÜRÜN TANIMA SİSTEMİ	47
3.1	Önerilen Yöntem.....	47
3.1.1	Ürün Tespiti Aşaması	49
3.1.2	Ürün Sınıflandırma Aşaması	54
3.1.3	İyileştirme Aşaması	59
3.2	Kullanılan Veri Setleri	61
4	DENEYSEL SONUÇLAR	65
4.1	GP-20 Veri Setinde Alınan Sonuçlar	65
4.1.1	Aşama-1 için Önerilen Yöntemlerin Performansı	65
4.1.2	Aşama-2 için Önerilen Yöntemlerin Performansı	66
4.1.3	Aşama-3 için Önerilen Yöntemlerin Performansı	69
4.1.4	Çok Aşamalı Uçtan-uca Ürün Tanıma Yaklaşımı için Kullanılan Tüm Yöntemlerin Performansı	69
4.2	GP-181 Veri Setinde Alınan Sonuçlar	73
4.3	Grocery Products Veri Setinde Alınan Sonuçlar	74
4.4	Çalışma Zamanı Performansı Değerlendirmesi	76
5	SONUÇ	78
	KAYNAKÇA	80
	TEZDEN ÜRETİLMİŞ YAYINLAR	87

SİMGE LİSTESİ

C'_l	l Sınıfı Olarak Tahmin Edilen Ürünün Sınırlayıcı Kutusu
k_l	l Sınıfında Doğru Olarak Sınıflandırılan Görsellerin Sayısı
C_l	l Sınıfındaki Etiketli Ürünün Temel Gerçek Sınırlayıcı Kutusu
n_l	l Sınıfındaki Toplam Görsel Sayısı
T	Toplam Sınıf Sayısı
x_2	Ürün Sınırlayıcı Kutusunun Sağ-alt Köşesinin Uzamsal x Koordinatı
x_{br}	Ürün Sınırlayıcı Kutusunun Sağ-alt Köşesinin Uzamsal x Koordinatı
y_2	Ürün Sınırlayıcı Kutusunun Sağ-alt Köşesinin Uzamsal y Koordinatı
y_{br}	Ürün Sınırlayıcı Kutusunun Sağ-alt Köşesinin Uzamsal y Koordinatı
x_{tr}	Ürün Sınırlayıcı Kutusunun Sağ-üst Köşesinin Uzamsal x Koordinatı
y_{tr}	Ürün Sınırlayıcı Kutusunun Sağ-üst Köşesinin Uzamsal y Koordinatı
x_{bl}	Ürün Sınırlayıcı Kutusunun Sol-alt Köşesinin Uzamsal x Koordinatı
y_{bl}	Ürün Sınırlayıcı Kutusunun Sol-alt Köşesinin Uzamsal y Koordinatı
x_1	Ürün Sınırlayıcı Kutusunun Sol-üst Köşesinin Uzamsal x Koordinatı
x_{tl}	Ürün Sınırlayıcı Kutusunun Sol-üst Köşesinin Uzamsal x Koordinatı
y_1	Ürün Sınırlayıcı Kutusunun Sol-üst Köşesinin Uzamsal y Koordinatı
y_{tl}	Ürün Sınırlayıcı Kutusunun Sol-üst Köşesinin Uzamsal y Koordinatı

KISALTMA LİSTESİ

ACF	Toplu Kanal Özellikleri (Aggregate Channel Features)
AGAST	Uyarlanabilir ve Genel Hızlandırılmış Segment Testi (Adaptive and Generic Accelerated Segment Test)
BES	Kel Kartal Arama (Bald Eagle Search)
BoW	Kelime Çantası (Bag of Words)
BRIEF	İkili Sağlam Bağımsız Temel Öznitelikler (Binary Robust Independent Elementary Feature)
BRISK	İkili Gürbüz Değişmeyen Ölçeklenebilir Kilit Noktaları (Binary Robust Invariant Scalable Key point)
CA	Kategorileştirme Doğruluğu (Categorization Accuracy)
CAlg	Kümeleme Algoritması (Clustering Algorithm)
CHM	Renk Histogramı Eşleştirme (Color Histogram Matching)
CNN	Evrişimli Sinir Ağı (Convolutional Neural Network)
DBN	Derin İnanç Ağı (Deep Belief Network)
DNN	Derin Sinir Ağı (Deep Neural Network)
DRN	Dinamik İyileştirme Ağı (Dynamic Refinement Network)
ELM	Aşırı Öğrenme Makineleri (Extreme Learning Machines)
EM-Merger	Beklenti-Maksimizasyon Birleşmesi (Expectation-Maximization Merger)
FAST	Hızlandırılmış Segment Testinden Özellikler (Features from Accelerated Segment Test)
FN	Yanlış Negatif (False Negative)
FP	Yanlış Pozitif (False Positive)
FPN	Özellik Piramit Ağı (Feature Pyramid Network)
GDN	Gaussian Kod Çözücü Ağı (Gaussian Decoder Network)
GLN	Gaussian Katman Ağı (Gaussian Layer Network)
HOG	Yönelimli Eğimlerin Histogramı (Histogram of Oriented Gradients)
HMM	Saklı Markov Modelleri (Hidden Markov Models)
HSV	Renk özü, Doygunluk, Parlaklık (Hue, Saturation, Value)
IoU	Kesiştirilmiş Bölgeler (Intersection over Union)
K-NN	K-En Yakın Komşu (K-Nearest Neighbour)
LCA	Yerel Kavramlar-Birikim (Local Concepts-Accumulation)

mAP	Genel Ortalama Hassasiyet (mean Average Precision)
MSE	Ortalama Kare Hatası (Mean Squared Error)
MSERF	Maksimum Kararlı Ekstrem Bölgeleri Özellikleri (Maximally Stable External Region Features)
NMS	Maksimum Olmayan Bastırma (Non-Maximum Suppression)
NR-CAIlg	Komşulukla İlişkili Kümeleme Algoritması (Neighborhood-Related Clustering Algorithm)
OCR	Optik Karakter Tanıma (Optical Character Recognition)
ORB	Yönelimli FAST ve Döndürülmüş BRIEF (Oriented FAST and Rotated BRIEF)
PA	Ürün Doğruluğu (Product Accuracy)
PAN	Yol Toplama Ağı (Path Aggregation Network)
pbl	Sol Alt Nokta (Point of Bottom Left)
pbr	Sağ Alt Nokta (Point of Bottom Right)
PP	Ürün Kesinliği (Product Precision)
PR	Ürün Duyarlılığı (Product Recall)
ptl	Sol Üst Nokta (Point of Top Left)
ptr	Sağ Üst Nokta (Point of Top Right)
R-CNN	Bölge Tabanlı Evrişimli Sinir Ağı (Region with Convolutional Neural Network)
ResNet	Artık Ağ (Residual Network)
ReLU	Rektifiye Doğrusal Birim (Rectified Linear Unit)
RFID	Radyo Frekansıyla Tanımlama (Radio Frequency Identification)
ROI	İlgilenilen Alan (Region of Interest)
RPN	Bölge Teklif Ağı (Region Proposal Network)
SCATTER	Seçici Bağlam Dikkatli Metin Tanıyıcı (Selective Context Attentional Text Recognizer)
SIFT	Ölçekten Bağımsız Öznitelik Dönüşümü (Scale-Invariant Feature Transform)
SKU	Stok Tutma Birimi (Stock Keeping Unit)
SNN	Siyam Sinir Ağı (Siamese Neural Network)
SPP	Uzamsal Piramit (Havuzu Spatial Pyramid Pooling)
SSD	Tek Atışta Çoklu Kutu Algılama (Single Shot MultiBox Detector)
SURF	Hızlandırılmış Dayanıklı Öznitelikler (Speeded-up Robust Features)
SVM	Destek Vektör Makineleri (Support Vector Machines)
TP	Gerçek Pozitif (True Positive)

YOLO	Sadece Bir Kez Bakarsın (You Only Look Once)
YOLOv2	Sadece Bir Kez Bakarsın versiyon 2 (You Only Look Once version 2)
YOLOv3	Sadece Bir Kez Bakarsın versiyon 3 (You Only Look Once version 3)
YOLOv4	Sadece Bir Kez Bakarsın versiyon 4 (You Only Look Once version 4)
YOLOv5	Sadece Bir Kez Bakarsın versiyon 5 (You Only Look Once version 5)



ŞEKİL LİSTESİ

Şekil 1.1 Market raflarında ürün tanıma probleminin zorluklarını temsil eden örnek görseller: (a, b) SKU-110K veri setinden farklı görüş açısına sahip raf görüntüleri [5], (c) Grocery Products veri setinden bulanık raf görüntüsü [6], (d, e) SKU-110K veri setinden farklı raf tasarımı görselleri [5], (f) SKU-110K veri setinden karmaşık arka plan raf görüntüsü [5], (g, h) Grocery Products veri setinde ambalaj tasarımı benzerliği yüksek olan farklı ürün çeşidi görselleri [6]	3
Şekil 1.2 Market raflarında ürün tanıma problemi ile ilgili yapılan çalışmaların anahtar kelimelere ve yıllara göre dağılımı.....	6
Şekil 2.1 Ürün tanıma problemi: Freiburg Groceries [46] veri setinden sınıflandırma örneği, SKU-110K [5] veri setinden ürün konumlarının tespiti örneği, Grocery Products [6] veri setinden ürün konumlarının tespiti ve sınıflandırılması örneği.....	21
Şekil 2.2 Farklı veri kümelerindeki farklı sınırlayıcı kutu bilgisi gösterim şekilleri: (a) Grozi-120 [31] veri setinden sınırlayıcı kutu örneği, (b) Grocery Products [6] veri setinden sınırlayıcı kutu örneği, (c) Unitail-Det'in [49] veri setinden sınırlayıcı kutu örneği.....	23
Şekil 3.1 Çok aşamalı uçtan-uca ürün tanıma süreci.....	48
Şekil 3.2 Aşama-1'de önerilen geleneksel bilgisayarla görme yaklaşımının adımları: (a) orijinal raf görüntüsü; (b) ilk adımdan sonra tespit edilen raf çizgilerini içeren raf görüntüsü; (c) ikinci adımdan sonra tespit edilen ürün bölgelerini içeren raf görüntüsü; (d) üçüncü adımdan sonra tamamlanmış ürün bölgelerini gösteren raf görüntüsü	50
Şekil 3.3 Aşama-1'de önerilen ACF detektörünün adımları.....	52
Şekil 3.4 Aşama-1'de önerilen YOLOv4 detektörünün adımları	52
Şekil 3.5 Aşama-1'de önerilen SSD adımları.....	53
Şekil 3.6 Aşama-2'de önerilen ön-eleme işleminin sözde kodu	55
Şekil 3.7 Aşama-2'de önerilen ürün sınıflandırma aşamaları	57
Şekil 3.8 Kullanılan veri setlerinden örnek görüntüler: (a,b) SKU-110K'nın örnek raf görüntüleri [5]; (c,d) Grocery Dataset'in örnek raf görüntüleri [17]; (e,f) GP-20'nin örnek raf görüntüleri [15]; (g,h) GP-181'in örnek raf görüntüleri [16]; (i,j) Grocery Products'ın örnek raf görselleri [6]; (k,l) Grocery Products ürünlerinin örnek şablon görüntüleri [6]	63

TABLO LİSTESİ

Tablo 1.1	Literatürdeki market raflarında ürün tanıma çalışmalarının sınıflandırılması	15
Tablo 1.1	Literatürdeki market raflarında ürün tanıma çalışmalarının sınıflandırılması (devamı).....	16
Tablo 2.1	Ürün tanımanın ele aldığı problemler	24
Tablo 2.2	Market ürünleri tanıma veri setlerine ait örnek görüntüler	28
Tablo 2.2	Market ürünleri tanıma veri setlerine ait örnek görüntüler (devamı) ...	29
Tablo 2.3	Market ürünleri tanıma veri setleri	30
Tablo 2.3	Market ürünleri tanıma veri setleri (devamı)	31
Tablo 2.4	Ürün tespiti için kullanılan yöntemler	35
Tablo 2.5	Özellik çıkarma yöntemleri	37
Tablo 2.6	Geleneksel özellik çıkarma yöntemleri.....	39
Tablo 2.7	Derin öğrenme tabanlı özellik çıkarma yöntemleri	42
Tablo 2.8	Özelliklerin sınıflandırılması için kullanılan yöntemler	44
Tablo 2.9	Sonuçların iyileştirilmesi için kullanılan yöntemler	46
Tablo 3.1	Farklı özellik çıkarma yöntemleri ile GP-20 veri setindeki ürün şablon görüntülerinden elde edilen özellik sayıları.....	58
Tablo 3.2	Kullanılan veri setlerinin özellikleri	62
Tablo 4.1	Aşama-1 için önerilen yöntemlerin GP-20 veri setindeki sonuçları.....	66
Tablo 4.2	SURF, BRISK ve ORB özelliklerinin GP-20 veri setindeki sonuçları.....	68
Tablo 4.3	Aşama-2 için önerilen yöntemlerin GP-20 veri setindeki sonuçları.....	69
Tablo 4.4	Aşama-3 için önerilen yöntemlerin GP-20 veri setindeki sonuçları.....	69
Tablo 4.5	Çok aşamalı uçtan-uca ürün tanıma yaklaşımı için kullanılan tüm yöntemlerin GP-20 veri setindeki sonuçları	70
Tablo 4.6	Önerilen yöntemin (SSD + SURF&BRISK&ORB + NR-CAlg) GP-20 veri seti üzerindeki karışıklık matrisi analizi	71
Tablo 4.7	Önerilen yöntemin (SSD + SURF&BRISK&ORB + NR-CAlg) GP-20 veri seti üzerindeki performansının yapılan çalışmalarla karşılaştırılması	72
Tablo 4.8	Önerilen yöntemin (SSD + SURF&BRISK&ORB + NR-CAlg) GP-181 veri seti üzerindeki performansının yapılan çalışmalarla karşılaştırılması	74
Tablo 4.9	Önerilen yöntemin (SSD + SURF&BRISK&ORB + NR-CAlg) Grocery Products veri seti üzerindeki performansının yapılan çalışmalarla karşılaştırılması	76

Tablo 4.10 Ürün sınıflandırma aşamasında kullanılan yöntemlerin çalışma sürelerinin GP-20 veri seti üzerinde karşılaştırılması.....	77
Tablo 4.11 Ürün sınıflandırma aşamasında kullanılan yöntemlerin hesaplama karmaşıklığı.....	77



Bilgisayarla Görme ve Makine Öğrenmesi Yöntemleri Kullanılarak Market Raflarında Ürün Tanıma

Ceren Gülra MELEK

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Programı

Doktora Tezi

Danışman: Prof. Dr. Songül VARLI

Eş-Danışman: Doç. Dr. Elena BATTİNİ SÖNMEZ

Market raflarındaki ürünlerin görüntüden tanınması için akıllı bir sistemin kullanılması, perakendecilik sektöründe karşılaşılan planogram uyumluluğu, stok takibi ve müşteri desteği gibi birçok sorunun çözümüne fayda sağlar. Ürün tanıma probleminin normal bir nesne tanıma probleminden farklı olarak kendine has zorlukları vardır: (1) ürün çeşitliliğinin ve ürünler arasındaki benzerliklerin fazla olması, (2) ürün boyutlarının birbirinden çok farklı olması (3) yenilenen ambalajlar ve eklenen yeni ürünlere ait verilerin sürekli toplanması ve sisteme eklenmesi. Bu nedenle, problemin çözümü için geleneksel yöntemlerin tek başına kullanılması yeterli değildir. Bu tez çalışmasında önerilen çok aşamalı uçtan-uca süreç ile ürün çeşidinden bağımsız olarak ürünlerin algılanması ve farklı ölçek, döndürme ve afin değişmez özellik çıkarma yöntemlerinin hibrit kullanımı sayesinde market raflarında ürün tanıma probleminin zorluklarına çözüm üretilmiştir. Ürünün algılanması, sınıflandırılması ve sonuçların iyileştirilmesini içeren bu yöntem ile girdi olarak alınan bir raf görüntüsünde önceden tanımlanmış tüm ürünlerin lokalizasyon ve sınıflandırma sonuçları elde edilir. Ürün algılama aşamasında,

görüntüdeki her bir mevcut ürünün lokalizasyonu SSD algoritması ile tespit edilir. Ürün sınıflandırma aşamasında, probleme uygun özellik çıkarma yöntemleri SURF, BRISK ve ORB'nin hibrit kullanımı ile ürün şablonları, ürün algılama aşamasından elde edilen bölge önerileriyle eşleştirilir. Son aşama olan iyileştirme aşamasındaysa, komşuluk ilişkisi eklenmiş kümeleme algoritması ile aynı ürüne ait çoklu bölge önerileri üzerinde iyileştirmeler yapılarak nihai sonuca ulaşılır. Geliştirilen yöntem, farklı sayıda ürün çeşidi ve görüntüden oluşan Grocery Products veri seti ve onun alt kümeleri olan GP-20, GP-181 veri setlerinde test edilmiştir. Ayrıca, ürün algılama aşamasında eğitim için Grocery Dataset ve SKU-110K'nın alt kümesinden bir eğitim seti oluşturulmuştur.

Bu tez çalışmasında, ürün çeşidinden bağımsız bir algılama süreci sağlandığından sürekli yenilenen ürün ambalajları ve eklenen yeni ürünlerden dolayı eğitim sürecinin tekrarlanması gerekmez. Ayrıca, hibrit bir ürün tanıma konseptiyle geniş ürün çeşitliliğinin market raflarındaki ürünlerin görüntüden tanınması probleminde yarattığı zorluğun üstesinden gelinmektedir. İyileştirme prosedürü ile güçlendirilen yöntemimiz, sunulan farklı yöntemlerle karşılaştırılır sonuçlar elde etmiştir.

Anahtar Kelimeler: Ürün tanıma, planogram uyumluluğu, stok takibi, tek atışta çoklu kutu algılama, özellik çıkarma.

Product Recognition on Grocery Shelves Using Computer Vision and Machine Learning Methods

Ceren Gülra MELEK

Department of Computer Engineering

Doctor of Philosophy Thesis

Supervisor: Prof. Dr. Songül VARLI

Co-supervisor: Assoc. Prof. Dr. Elena BATTİNİ SÖNMEZ

Using an intelligent system to recognize products from grocery shelf images provide to solve many problems encountered in the retail industry, such as planogram compatibility, stock tracking and customer support. The product recognition problem, unlike a normal object recognition problem, has its own difficulties: (1) product diversity and similarities between products are high, (2) product sizes are very different from each other (3) constantly collecting and adding the data of renewed packaging and new products to the system. Therefore, the use of conventional methods alone is not enough to solve problem. In this thesis, the proposed multi-stage end-to-end process produces solution to the difficulties of product recognition problem on grocery shelves with a product-independent detection process and the hybrid use of different scale, rotation, and affine invariant feature extraction methods. The method which includes product detection, classification and refinement of results, returns the localization and label results of all predefined products in a shelf image that takes as input. In the detection phase, the localization of each existing product in the image is detected with SSD algorithm. In the classification phase, product templates are matched with the region proposals obtained from the detection phase with the hybrid use of problem-

appropriate feature extraction methods: SURF, BRISK and ORB. In refinement, the final result is achieved by making improvements on multiple region proposals for the same product with the neighborhood-related clustering algorithm. The developed method was tested on the dataset of Grocery Products and its subsets GP-20 and GP-181, which consists of different numbers of product types and images. Additionally, a training set was created from the Grocery Dataset and subset of SKU-110K for training in detection phase.

In this thesis, retraining process does not required due to constantly renewed packaging and newly added products because a product-independent detection is provided. Moreover, a hybrid product recognition concept overcomes the difficulty created by product diversity in the problem of product recognition from shelf images. Our method, enhanced by the refinement procedure, achieved comparable results with the different presented methods.

Keywords: Product recognition, planogram compliance, stock-tracking, single shot multibox detector, feature extraction.

1.1 Tezin Amacı

Teknolojik gelişmeler ve endüstriyel dönüşüm, insanların günlük yaşamdaki sıradan iş faaliyetlerinde bile beklentilerinin her geçen gün artmasına neden olmaktadır. Bu beklentileri karşılamak amacıyla, sağlıktan ulaştırmaya, sanayiden enerjiye, eğitimden güvenliğe kadar farklı birçok alandaki gerçek dünya problemlerine bilgisayarla görme ve makine öğrenmesi yöntemleri kullanılarak uygun çözümler bulunmaktadır. Perakende sektöründe karşılaşılan planogram uyumluluğu, stok takibi ve müşteri desteği gibi önemli problemlerin çözümü için de market raflarındaki ürünlerin bilgisayarla görme ve makine öğrenmesi yöntemleri kullanılarak görüntüden tanınması mümkündür.

Perakende sektöründe, tüketiciyi etkileyerek satın almaya yönlendirmek amacıyla mağaza içinde uygulanan her türlü aktivite, teknik veya metot tanzim-teşhir olarak tanımlanmaktadır. Tanzim-teşhir, ürünlerin market içindeki bulunabilirliğini, görünebilirliğini ve satın alınabilirliğini arttırmak için belirlenen kurallar çerçevesinde, ürünlerin en uygun yerde ve müşterilerin ilgisini en fazla çekecek şekilde sergilenmesi faaliyetlerini içermektedir [1]. Tanzim-teşhir uygulamalarının önemli bir parçası olan planogramlar ile perakende ürünlerin satışı en arttırıcı şekilde market raflarındaki ideal düzenlemesini belirten şemalar oluşturulur [1]. Bu şemaların raflardaki gerçek yerleşim ile uyumluluğunun sürekli olarak kontrol edilmesi ve ürün sayılarının takip edilerek eksilen ürünlerin tamamlanması gerekmektedir. Günümüzde bu işlemlerin gerçekleştirilmesi için izlenen genel prosedür, görevli kişiler tarafından market raflarının fotoğraflarının belli aralıklarla çekilmesi ile manuel olarak yapılmaktadır. Ancak, manuel olarak yapılan bu işlem insan hatasına açıktır, zaman kaybı yaratır ve fazladan insan gücü maliyeti oluşturur.

Otomatik bir ürün tanıma sisteminin temel rolü, marketlerden elde edilen raf görüntülerindeki ürünlerin konumunu ve sınıfını bulmaktır. Böyle bir ürün tanıma sisteminin varlığı hem üreticilere, hem tedarikçilere ve hem de müşterilere çok sayıda fayda sağlar. Üreticiler, ürünlerinin market raflarında planograma uygun bir

şekilde yerleştirilmesini ister. Yapılan bir araştırmaya göre, ürünler optimize edilmiş planogram ile mükemmel bir şekilde eşleştğinde iki hafta içinde satışlar %7,8, kar oranı ise %8,1 artmıştır [2]. Dolayısıyla, planogramların oluşturulması kadar planograma uygun ürün yerleştirilmesinin yapılıp yapılmadığının kontrolü de oldukça önemlidir. Otomatik bir ürün tanıma sistemi ile üreticiler, bu kontrolü hızlı ve kolay bir şekilde gerçekleştirebilir. Tedarikçiler tarafındaysa otomatik bir ürün tanıma sisteminin varlığı, stok takibi, envanterin yenilenmesi ve planogram uyumluluğu açısından faydalıdır. Özellikle büyük marketlerin yönetimi için gerekli bilgiler gerçek zamanlı olarak sağlanır. Otomatik ürün tanıma sisteminden elde edilen bilgiler, stokta olmayan veya yanlış yere yerleştirilen ürünleri tespit ederek satışları artırmayı ve müşterilerin alışveriş kalıplarını analiz ederek müşteri deneyimini iyileştirmeyi mümkün kılmaktadır. Ayrıca, stokta olmayan ürünlerle ilgili yapılan araştırma [3], rafta istediği ürünü bulamayan müşterilerin %31'inin aynı ürünü başka bir marketten aldığını, %22'sinin aynı ürünün farklı markasını satın aldığını ve %11'inin ürünü bulamayınca almaktan vazgeçtiğini göstermektedir. Bu nedenle, raflardaki ürün sayısının takip edilerek eksilen ürünlerin tamamlanması ile tüketicinin stokta var olan ürünü rafta bulamamasının önüne geçmek ve yine planogram uyumluluğunu devam ettirmek gerekmektedir. Müşteri tarafında, başarılı bir ürün tanıma sistemi geliştirilmiş bir müşteri deneyimi sağlar. Müşteriler aradıkları ürünü daha hızlı ve daha kolay bir şekilde bulabilir ve fiyat karşılaştırmaları yapabilir. Görme engelli müşteriler için de daha konforlu bir alışveriş deneyimi sunar.

Tüm paydaşlar açısından değerlendirildiğinde, market raflarındaki ürünlerin görüntüden tanınması hem zamandan tasarruf sağlayan, hem de satışları [2] ve müşteri memnuniyetini artıran [4] bir çözüm oluşturmaktadır. Ancak, Şekil 1.1'de gösterildiği üzere otomatik bir ürün tanıma sisteminin, yaygın olarak karşılaşılan bilgisayarla görme zorluklarının (gereken sahne karmaşıklığı, bulanıklık, düzensiz aydınlatma koşulları, farklı görüş açısı vb.) yanı sıra, problemin kendine has zorluklarının da üstesinden gelmesi beklenmektedir. Market raflarında yer alan ürünlerin farklı raf tasarımlarında sergilenmesi, farklı ürün boyutlarının ve sürekli değişen ürün ambalajlarının olması, farklı ürün türleri arasında ambalaj tasarımı benzerliğinin yüksek olması ve ürün sınıflarının fazla olması ürün tanıma probleminin kendine has zorlularındandır.



Şekil 1.1 Market raflarında ürün tanıma probleminin zorluklarını temsil eden örnek görseller: (a, b) SKU-110K veri setinden farklı görüş açısına sahip raf görüntüleri [5], (c) Grocery Products veri setinden bulanık raf görüntüsü [6], (d, e) SKU-110K veri setinden farklı raf tasarımı görselleri [5], (f) SKU-110K veri setinden karmaşık arka plan raf görüntüsü [5], (g, h) Grocery Products veri setinde ambalaj tasarımı benzerliği yüksek olan farklı ürün çeşidi görselleri [6]

Bu tezde, market raf görsellerindeki ürünlerin tanınması sorunu ele alınmaktadır. Bu amaçla, girdi olarak alınan bir raf görüntüsünde önceden tanımlanmış tüm ürünlerin konumları ve sınıf bilgileri elde edilir. Önerilen çok aşamalı süreç, yukarıda belirtilen zorlukları, üründen bağımsız bir ürün algılama süreci ve farklı ölçek, döndürme ve afin değişmez özellik çıkarma yöntemlerini bir arada kullanarak ele alır. Önerilen yöntem üç ana adımdan oluşmaktadır. Ürün algılama aşaması, görüntüdeki her mevcut ürünün lokalizasyonunu dört farklı yöntemle bulur: önerilen geleneksel bilgisayarla görme yaklaşımı, toplu kanal özellikleri (ACF, Aggregate Channel Features) [7], sadece bir kez bakarsın versiyon 4 (YOLOv4, You Only Look Once version 4) [8] ve tek atışta çoklu kutu algılama (SSD, Single Shot MultiBox Detector) [9]. Ürün sınıflandırma aşaması, ürün şablonlarını ve ürün algılama sonuçlarını, hızlandırılmış dayanıklı öznitelikler (SURF, Speed-up Robust Features) [10], ikili gürbüz değişmeyen ölçeklenebilir kilit noktaları (BRISK,

Binary Robust Invariant Scalable Key point) [11], yönelimli hızlandırılmış segment testinden özellikler (FAST, Features from Accelerated Segment Test) [12] ve döndürülmüş ikili sağlam bağımsız temel öznitelikler (BRIEF, Binary Robust Independent Elementary Feature) [13] (ORB, Oriented FAST and Rotated BRIEF) [14] gibi uygun özellik çıkarma yöntemleri ve bunların hibrit kullanımı ile eşleştirir. Üçüncü aşama olan iyileştirme aşamasında, sınıflandırma ve lokalizasyon sonuçlarını komşuluk ilişkisi eklenmiş kümeleme algoritması ile iyileştirerek nihai sonuca ulaşılır. Grocery Products [6] veri kümesinin tamamı ve alt kümeleri (GP-20 [15], GP-181 [16]) önerilen yöntemleri test etmek için kullanılır. Ek olarak, Grocery Dataset [17] ve SKU-110K'nın [5] alt kümesinden bir eğitim seti oluşturulmuş, ACF [7], YOLOv4 [8] ve SSD'nin [9] eğitimi için kullanılmıştır.

1.2 Özgün Değer

Bu tez çalışmasının özgün değeri aşağıdaki üç madde ile özetlenmiştir:

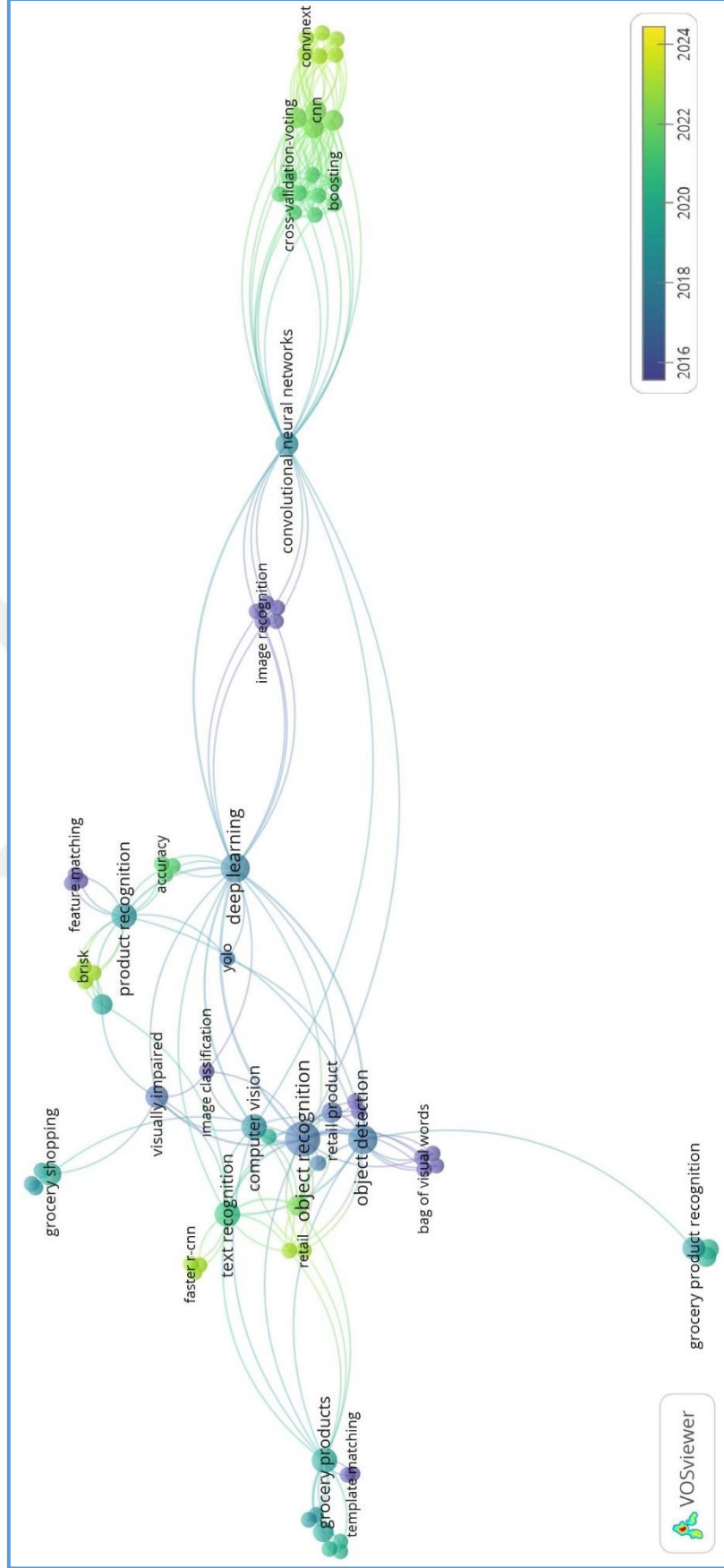
1. Market raflarında ürün tanıma probleminin en önemli zorluklarından biri olan ürün ambalajlarının sürekli değişmesi ve market ürünlerine yeni ürün çeşitlerinin eklenmesi otomatik ürün tanıma sistemlerinin yeniden eğitimi gerektirmektedir. Önerilen üründen bağımsız algılama süreci, eklenen yeni sınıflar için ürün tespiti sırasında yeniden eğitim gerektirmeyen bir yaklaşım sunar.
2. Market raflarındaki ürün çeşitliliğinin fazla olması ve bu ürünlerin özelliklerinin birbirinden çok farklı olması sebebi ile farklı yöntemler farklı ürün çeşitlerinde daha iyi sonuçlar verebilmektedir. Önerilen farklı özellik çıkarım yöntemlerinin bir arada kullanıldığı hibrit bir model ile ürün tanıma probleminin bu zorluğuna bir çözüm sunulmuştur.
3. Önerilen çok aşamalı yaklaşımın ürün tanımadaki performansı, ürün algılama ve sınıflandırma aşamalarında sunulan farklı bilgisayarla görme ve makine öğrenmesi yöntemleriyle ve literatürdeki başlıca çalışmalarla karşılaştırılmıştır. Ayrıca, önerilen sistemin performans ve çalışma süresinin dengeli olduğunu göstermek amacıyla çalışma zamanı değerlendirmesine de yer verilmiştir.

1.3 Literatür Taraması

Market raf görüntülerinde ürün tanıma, sağladığı çeşitli faydalar nedeniyle hem akademik hem de endüstriyel açıdan önemli bir araştırma alanı haline gelmiştir. Planogram eşleştirme, ürün tanıma ve stok takibi gibi market içi sorunları çözmek için literatürde çeşitli çözümler önerilmiştir. Yaygın olarak kullanılan çözümler arasında barkod [18], [19] ve radyo frekansı tanımlama (RFID, Radio Frequency Identification) etiketleri [20]–[23] ile bilgisayarla görme ve makine öğrenmesi yöntemleri yer almaktadır.

Barkodlar, ürün yüzeyinde yer alan çizgilerden oluşan ve sayısal veya alfa-numerik bir kodu temsil eden benzersiz grafik desenleri ile ürünlerin tanımlanmasını ve takip edilmesini sağlar [24]. Ürün bilgilerini okumak ve ürünü tanımlamak için özel bir tarayıcı veya kamera gerektirir. RFID ise cihazlar arasında veri iletmek için radyo frekansı bilgisi içeren etiketleri kullanarak ürünlerin tanımlanmasına olanak sağlar [25]. Barkod ve RFID teknolojilerinin kullanımının birçok uygulamada ürünlerin tanınmasında etkili olduğu kanıtlanmıştır ancak bazı sınırlamaları da mevcuttur. Barkod tanıma için barkodun yer aldığı yüzeyin tarayıcının ya da kameranın görüş açısında yer alması gerekir. Ayrıca barkodlar, çevresel zararlara açık olmasıyla birlikte ürün tanımada gecikmelere veya hatalara neden olmaktadır [26]. RFID tarafında ise donanım gerekliliği, çevre koşullarına duyarlılık, veri güvenliği ve bilgi çarpışmaları gibi sorunlarla karşılaşmaktadır [27].

Günümüzde, yeni yaklaşımlara duyulan ihtiyaçla birlikte veri toplama kolaylığının artması, veri depolama ve işleme maliyetlerinin azalması, bu tür donanım tabanlı çözümler yerine bilgisayarla görme ve makine öğrenmesi yöntemlerinin kullanımının artmasına neden olmuştur. Makine öğrenmesi algoritmaları, yeni öğeleri tanıyacak ve değişen ortamlara uyum sağlayacak şekilde eğitime olanak sağlar. Bu durum, ürün tanıma konusunda barkod veya RFID gibi geleneksel teknolojilere göre makine öğrenmesini daha esnek bir seçenek haline getirir. Bilgisayarla görme algoritmaları da görüntünün renksel, yapısal, morfolojik ve frekans gibi özelliklerini analiz ederek ürünleri yüksek doğrulukla tanımlayabilmektedir. Bu nedenlerle, market raf görüntülerindeki ürünlerin tanınması için bilgisayarla görme ve makine öğrenmesi yöntemleri kullanılarak yapılan çalışmaların sayısı artmaktadır.



Şekil 1.2 Market raflarında ürün tanıma problemi ile ilgili yapılan çalışmaların anahtar kelimelere ve yıllara göre dağılımı

Market raflarında ürün tanıma problemi ile ilgili yapılan çalışmaların anahtar kelimelere ve yıllara göre dağılımı Şekil 1.2’de verilmiştir. Bu dağılıma göre ürün tanıma alanında yapılan çalışmalar 2016 yılından günümüze kadar uzanmaktadır. Yapılan çalışmaların çoğunda nesne tanıma, nesne tespiti ve derin öğrenme anahtar kelimelerinin öne çıktığı görülmektedir. Kullanılan yöntemler açısından incelediğimizde de birçok bilgisayarla görme ve makine öğrenmesi yönteminin öne çıktığı görülmektedir. Bunlar arasında özellik eşleştirme, şablon eşleştirme, BRISK [11], kelime çantası (BoW, Bag of Word), daha hızlı bölge tabanlı evrişimli sinir ağı (Faster R-CNN, Faster Region with Convolutional Neural Network) [28], sadece bir kez bakarsın (YOLO, You Only Look Once) [29], derin öğrenme, boosting algoritması, evrişimli sinir ağı (CNN, Convolutional Neural Network) ve ConvNeXt [30] gibi terimler öne çıkmaktadır. Son yıllarda yapılan çalışmalara bakıldığında da derin modellerin öne çıktığı gözükmemektedir.

Baştan sona ele alınan bir market raflarında ürün tanıma probleminde girdi olarak alınan bir raf görüntüsünde önceden tanımlanmış tüm ürünlerin lokalizasyon ve sınıflandırma sonuçları elde edilir. Aynı zamanda tüm bu süreç, görüntüdeki rafların tespiti, ürün konumlarının tespiti, ürünün sınıflandırılması gibi farklı alt problemlerden oluşur. Yapılan çalışmalar incelendiğinde, Tablo 1.1’de görüldüğü üzere bu sorunu bir bütün olarak ele alan çalışmaların [6], [31]–[36] yanı sıra problemin farklı kısımlarına ayrı ayrı çözümler getiren [5], [17], [37]–[46] ve farklı aşamalardan elde edilen sonuçları birleştiren [15], [16], [47] çalışmalar mevcuttur. Ayrıca ürün tanıma problemini, görselin özelliklerinden tanımak yerine ürün üzerindeki metni tanıyarak çözmeye çalışan çalışmalar da mevcuttur [48]–[50]. Yapılan bu çalışmalardan farklı olarak sınıflandırma gerçekleştirmeksizin rafta birbiri ile aynı ürünleri bulmaya dayalı yaklaşımlarla da perakende sektörü problemlerine çözümler [51], [52] önerilmiştir.

Sınıflandırma yapmadan rafta birbiri ile aynı ürünleri bulan çalışmalarda eğitim kümesine gerek duyulmaz ama bu seferde sistemin farklı girdilere ihtiyacı vardır. Rafta birbiri ile aynı ürünleri bularak ürün sayısını hesaplayan çalışma [51]’de kullanıcının, raf görüntüsü üzerinde aranacak olan ürünün ilgilenilen alanını (ROI, Region of Interest) ve ROI içerisinden seçilen nesnenin en ayırt edici özelliğinin bulunduğu alt alanını seçmesi gerekmektedir. Önerilen sistem, seçilen üründen rafta kaç adet bulunduğunu hesaplamak için kullanıcıdan girdi olarak aldığı alanları raf

görüntüsündeki ürünler ile morfolojik özellikleri ve renk özü, doygunluk, parlaklık (HSV, Hue, Saturation, Value) renk histogramları üzerinden eşleştirir. Tekrarlanan örüntüleri kullanarak planogram uyumluluğu kontrolünün yapıldığı çalışma [52]'de, sistem girdi olarak raf sayısı, raf başına düşen ürün sayısı ve ürün türü gibi planogram bilgilerini alır. Girdi olarak alınan bilgilerden elde edilen alanlar ile raf görüntüsünden elde edilen tekrarlanan örüntüler arasında spektral grafik uyumu eşleştirmesi yapılarak planogram uyumluluğu kontrol edilir.

1.3.1 Üç Aşamalı Uçtan-uca Olmayan Ürün Tanıma Çalışmaları

Bir ürün tanıma sisteminin performansı, üç temel problemin çözümüne bağlıdır: raf tespiti, ürün tespiti ve ürün sınıflandırması. Raf görüntülerinde raf çizgilerinin tespiti, ürünü daha sınırlı bir alanda aramak ve yanlış alanlarda tespit edilen ürünleri elemek açısından performans artırıcı bir etkiye sahiptir. Önceki çalışmalar [17], [37] ve [39], Grocery Dataset [17] veri seti üzerinde raf tespiti sorununu ele almıştır. Raf çizgilerini tespit etmek için çalışma [37]'de, Hough dönüşümü [53] kullanılarak 229 raf görüntüsü üzerinde %83,4 doğruluk elde edilmiştir. Hough dönüşümü [53] raf tasarımıdaki değişikliklere karşı dayanıklı bulunmadığından ve hesaplama açısından daha karmaşık olduğundan, çalışma [17]'de raf çizgileri, ürünlerin y eksenindeki izdüşümünün histogramı hesaplanarak tespit edilmiştir. Çalışma [39]'da ise Gaussian karışım modeli kullanarak 350 raf görüntüsünde %99,03 doğruluk elde edilmiştir.

Yapılan çalışmalar [17], [37] ve [39]'un yazarları, ürün tespit sorununu farklı yöntemlerle ele almışlar ve performanslarını Grocery Dataset [17] veri setinin farklı kombinasyonlarında test etmişlerdir. Çalışmalarda ürün tespiti için yönelimli eğimlerin histogramı (HOG, Histogram of Oriented Gradients) [54] öznitelikleri kademeli nesne tanıma algoritması [55] ile sınıflandırılmıştır. Çalışma [39]'da buna ek olarak elde edilen yanlış sonuçlar Gauss karışım modeliyle çıkartılmıştır. Ayrıca, tespit edilemeyen ürün kutularını oluşturmak için de ortalama ve medyan filtre kullanılmıştır. Böylece en iyi performans %97,31 duyarlılık ve %94,05 kesinlik değerleri ile çalışma [39]'da elde edilmiştir.

Ürün sınıflandırma aşaması için de [17], [37] ve [39] çalışmalarında çeşitli çözümler önerilmiştir. Ürün tespiti ile elde edilen alanlardaki markaların tanınması için çalışma [17]'de ölçekten bağımsız öznitelik dönüşümü (SIFT, Scale-Invariant

Feature Transform) [56] ile HSV renk özellikleri, çalışma [37]'de ise HOG ve HSV renk özellikleri destek vektör makineleri (SVM, Support Vector Machines) kullanılarak sınıflandırılmıştır. Bunlardan farklı olarak çalışma [39]'da, elde edilen yoğunSIFT [56] özellik vektörlerinin aşırı öğrenme makineleri (ELM, Extreme Learning Machines) [57] kullanılarak sınıflandırılması ile %99,21 doğruluk elde edilmiştir.

Bu çalışmalar, ürün tanıma sorununun farklı aşamalarına çeşitli çözümler sunmuşlardır. Ancak, perakende alanındaki gerçek kullanım için uçtan-uca bir sistem performansı sunmamışlardır. Ürün tanıma sisteminin her aşaması için önerilen yöntemlerin performansları ayrı ayrı ele alınmıştır. Ayrıca, Grocery Dataset [17] veri seti yalnızca sigara paketlerinden oluşan sınırlı bir koleksiyondur. Önerilen yöntemler, ürün çeşitliliği açısından daha karmaşık veri setlerini (farklı ürün boyutları ve şekillerinden oluşan) test etmemiştir ve bu sistemlerin diğer veri setlerinde aynı performansı gösterebileceği konusunda bir bilgi bulunmamaktadır.

1.3.2 Ürün Tespit Aşamasını Ele Alan Çalışmalar

Çalışma [5], [38], [41] ve [43]'te, ürün çeşidine bakılmaksızın yalnızca rafta yer alan ürünlerin konumlarının tespit edilmesi problemine odaklanılmıştır. Literatürde yaygın olarak kullanılan nesne tespiti algoritmaları: daha hızlı bölge tabanlı evrişimli sinir ağı Faster R-CNN [28], YOLO9000 [29] algoritması, RetinaNet [58], sadece bir kez bakarsın versiyon 3 (YOLOv3, You Only Look Once version 3) [59] ve CenterNet [60] ile elde edilen sonuçlar, [5], [38], [41] ve [43]'te çalışmalarında geliştirilen yöntemlerle karşılaştırılmıştır.

Çalışma [5]'te yazarlar, çok fazla ürünün yer aldığı yoğun nesne tespiti için geliştirilmiş RetinaNet [58] ve beklenti-maksimizasyon birleşmesi (EM-Merger, Expectation-Maximization Merger) ile birleştirilmiş bir çerçeve önermişlerdir. Sistemin eğitimi ve testi için raflarda yer alan ürünlerin ürün çeşidine bakılmaksızın tek bir sınıf olarak etiketlendiği SKU-110K [5] veri seti kullanılmıştır. Bu çalışmada, [28], [29] ve [58]'de sunulan mevcut algoritmalarla karşılaştırıldığında geliştirilmiş bir performans elde edilmiştir. Çalışma [41]'de ise, RetinaNet'in [58] genişletilmiş bir versiyonu olan Gaussian kod çözücü ağı (GDN, Gaussian Decoder Network) ve GDN'ye kıyasla daha az parametreye sahip ve daha yüksek doğruluk değeri sunan Gaussian katman ağı (GLN, Gaussian Layer Network) adlı iki farklı

model önerilmiştir. Hem GDN hem de GLN, çalışma [5]'te önerilen yöntemden daha iyi bir performans göstermiştir. Ayrıca en yüksek performans, iki modülden oluşan (özellik seçimi ve dinamik iyileştirme başlıkları) dinamik iyileştirme ağı (DRN, Dynamic Refinement Network) ile çalışma [38]'de elde edilmiştir. Bir diğer çalışma [43]'te önerilen yöntem, Cascade R-CNN [61], Artık Ağ-101 (ResNet, Residual Network), özellik piramit ağı (FPN, Feature Pyramid Network) ve dengeli L1 kaybından oluşmaktadır. Yoğun sahnelerden kaynaklanan düşük kaliteli bölge önerileri, Cascade R-CNN [61] ile iyileştirilmiş, kayıp fonksiyonundan kaynaklanan konum kaybı ise dengeli L1 kaybı kullanılarak dengelenmiş ve sınırlandırılmıştır. Ayrıca, FPN ile küçük ölçekli nesnelerin tespiti üzerinde performans artışı sağlanmıştır. Çalışma sonucunda, Faster R-CNN [28], YOLOv3 [59] ve RetinaNet [58] gibi bilinen nesne tespiti yöntemlerine göre daha yüksek tespit doğruluğuna ulaşılmıştır. Ancak, kullanılan kademeli yapı nedeniyle önerilen yöntemin hızı yavaş olup, market raflarındaki ürünleri gerçek zamanlı olarak tespit etme gereksinimlerini karşılamamaktadır.

Raftaki ürünlerin konumlarının tespit edilmesi için çalışma [5], [38], [41] ve [43]'te karşılaştırılabilir önemli sonuçlar elde edilmiştir. Ancak bir ürün tanıma problemi için gerekli olan konum bilgilerine ulaşılırken, hangi konumda hangi ürün çeşidi bulunduğu bilgisi göz ardı edilmiştir. Dolayısıyla böyle bir sistem, raftaki ürün çeşidinin önemli olmadığı sadece ürün konumlarının önemli olduğu perakende çözümleri için uygundur. Bununla birlikte, SKU-110K [5] gibi yoğun ürün içeren veri setlerinde ürünlerin sınıflandırılması oldukça zorlu bir problemdir.

1.3.3 Ürün Sınıflandırma Aşamasını Ele Alan Çalışmalar

Çalışma [40], [42], [44]–[46] ve [50]'de ürün tanıma probleminin ürün sınıflandırılması aşamasına odaklanılmıştır. Çalışma [40]'da Instagram üzerinde önceden eğitilmiş evrişimli ağ ResNext-INet, yeni bir sinir ağı katmanı olan yerel kavramlar-birikim (LCA, Local Concepts-Accumulation) ve maksimum entropi kaybı ile ince ayar yapılarak, üç farklı market ürünü veri setini sınıflandırmak için kullanılmıştır. Elde edilen sonuçlara göre, önerilen yöntem, anahtar noktaların tespitine dayalı görüntü eşleştirme ve ince ayar yapılmayan ResNext-INet'e kıyasla daha yüksek bir doğruluk elde etmiştir. Bu çalışma, ürün sınıflandırması için evrişimli bir ağı kullanılmasının ve ağda yapılan ince ayarın sistemin doğruluğunu

arttırabileceğini göstermiştir. Bu durumda, her sınıf için birden fazla eğitim görüntüsüne sahip olmak, ağın eğitimi için önemlidir.

Bir diğer çalışma [42]'de ürün sınıflandırması için geliştirilmiş Siyam sinir ağı (SNN, Siamese Neural Network) [62] kullanılmıştır. Ağın geliştirmesi, çapraz-entropi kaybı fonksiyonunun, öklid uzaklığı ile güncellenmesi ve çift dikkat mekanizmaları eklenmesi suretiyle gerçekleştirilmiştir. Önerilen algoritma, yaygın olarak kullanılan iki veri setinde test edildiğinde daha geleneksel teknikleri geride bırakmıştır. Bu çalışma, [40]'da da karşılaşılan ve perakende ürün tanıma alanındaki önemli sorunlardan biri olan eğitim aşamasındaki yetersiz veri probleminin üstesinden gelmiştir. Öte yandan, market raflarındaki ürün tanıma sorununun sadece sınıflandırma aşamasına çözüm üretilmiş olup, baştan sona bir çözüm önerilmemiştir.

Diğer çalışmalardan farklı olarak [44]'te, görme engelli kişilerin alışveriş deneyimini iyileştirmek için seçilen ürünlerin tanınmasını sağlayacak alt metin oluşturulmaktadır. Öncelikle, giriş görselleri arasından alternatif metin olmayan görsellerin seçilmesi için derin inanç ağı (DBN, Deep Belief Network), seçilen görsellere alternatif metin üretilmesi için ise kel kartal arama (BES, Bald Eagle Search) algoritması kullanılmıştır. Önerilen teknik, daha düşük bir ortalama kare hatası (MSE, mean squared error) değerine sahipken, kesinlik, duyarlılık, doğruluk ve başarı oranı açısından mevcut sınıflandırıcılardan daha iyi bir performans göstermiştir. Sadece dijital ortamlar için geliştirilen bu çalışmada market ürünü tanıma probleminin daha spesifik bir kısmına odaklanılmıştır.

WebMarket isimli yeni bir veri setinin oluşturulduğu çalışma [45]'te amaç, verilen bir ürün şablonunu içeren raf görsellerinin bulunmasıdır. Raf görüntülerinde ilgilenilen bölgeleri tanımlamak için Harris-Affine [63] ilgi bölgesi detektörü kullanılmıştır. Daha sonra SIFT özelliklerinden hiyerarşik k-ortalamlar kümelemesi kullanılarak bir kelime dağarcığı oluşturulup, görsel kelimelerin histogramları kullanılarak ürün şablonlarının temsilleri oluşturulmuştur. Raf görselleri de alt görsellere bölünüp her bir alt görsele ait kelimeler, dört farklı benzerlik ölçüsü kullanılarak hedef ürün görselinin histogramıyla karşılaştırılmıştır. Önerilen yöntemin performansı bir gerçek hayat uygulaması için yeterli değildir.

Ürün tanıma veri seti Freiburg Groceries'in sunulduğu çalışma [46]'da, raf görüntülerinden manuel olarak kırılarak elde edilen ürün görüntüleri bir CNN modeli kullanılarak sınıflandırılmıştır. CNN modeli olarak, AlexNet'in [64] bir uyarlaması olan CaffeNet mimarisi [65] yeniden eğitilmiştir. Beş farklı eğitim test bölümünün ortalaması ile elde edilen sınıflandırma performansı %78,9'dur. Çalışma hem sunulan veri seti ile hem de derin ağlarla ürün sınıflandırılması açısından bir temel oluşturmuştur.

Müşterilerin alışveriş deneyimini iyileştirmeyi hedefleyen bir diğer çalışma [50]'de araştırmacılar ürünleri tanımak için metinsel ve görsel özelliklerin her ikisinden de yararlanmışlardır. Metin tanıma için ürün paketleri üzerindeki kelimeler tespit edilerek optik karakter tanıma (OCR, Optical Character Recognition) [66], görsel özellikler için de ayırt edici yamalar kullanılmıştır. Elde edilen özellikler, SVM kullanılarak sınıflandırılmıştır. Ayrıca, çalışmanın performansı kullanıcıdan alınan geri bildirimler yoluyla aktif öğrenme [67] kullanılarak geliştirilmiştir. Önerilen yöntem sadece ince taneli sınıfların ait olduğu kategorileri tahmin edebilmektedir. Daha detaylı sınıflandırma yapılabilmesi için sistemin geliştirilmesi gerekmektedir.

1.3.4 Ürün Tespiti ve Ürün Sınıflandırma Aşamalarının Ortak Olarak Gerçekleştiren Çalışmalar

Ürünlerin tespiti ve ürün sınıflandırılması sorununa ayrı ayrı odaklanan çalışmaların yanı sıra bir bütün olarak ele alan çalışmalar da [6], [31]–[36] bulunmaktadır. Çalışma [31]'de, raf görüntülerinden alınan pencere kareleri üzerinde beş farklı ölçek kullanılarak üç farklı yöntemle uçtan-uca ürün tanıma işleminin karşılaştırmalı sonuçları verilmiştir. Elde edilen kesinlik ve duyarlılık değerlerine göre en iyi performansa SIFT [56] ile ulaşılmış, bunu renk histogramı eşleştirme (CHM, Color Histogram Matching) ve son olarak da Boosted-Haar [55] özellikleri izlemiştir. Bu çalışma başarılı sonuçlar vermese de ürün tanıma probleminin çözümü için bir temel oluşturmaktadır.

Çalışma [32]'de yazarlar, görme engelli kullanıcılar için, alışveriş listelerinde yer alan ürünleri perakende raflarındaki video görüntülerinden bulmayı amaçlayan bir uygulama önermişlerdir. SURF [10] tanımlayıcıları ve çok sınıflı Naive Bayes sınıflandırıcısı [68], optik akışa dayalı bir yöntemle [69] elde edilen bölgelerden gelen ürünleri tanımak için kullanılmıştır. Sistemin performansı, 25 farklı alışveriş

listesinden 10 ürünün 10 farklı video görüntüsünde tespit edilmesiyle ölçülmüştür. Araştırma sonuçlarına göre doğru ürünü elde etme eşiği arttıkça kaçırılan ürün sayısı da artmıştır. Ayrıca, eşik değeri düştükçe de toplanan yanlış ürün sayısı artmıştır. Bu nedenle, çalışma kullanıcıların alışveriş listelerini tamamlayabilmeleri için uygun bir eşik değeri önerememiştir.

Çalışma [6]'da önerilen yöntem üç ana adımdan oluşmaktadır. Ürün tespiti ve ürün tanıma aşamalarının birlikte ele alındığı ilk iki adımda birbirini takip eden iki sıralama prosedürü yoluyla verilen bir test görüntüsünün ait olabileceği kategoriler filtrelenir ve filtrelenen kategorilerdeki tüm görüntülerle eşleştirilir. Bunun için yoğun SIFT özelliklerini [56] kullanan bir bölge-kısıtlamalı doğrusal kodlama modeli rastgele orman algoritması [70] ile eğitilmiştir. Ardından deforme olabilir uzamsal piramit eşleştirme yoluyla hızlı yoğun piksel eşleştirmesi yapılmıştır. Son adımda ise, tanınan ürünlerin çıkarımsal konumlarıyla birlikte nihai listesini elde etmek için bir enerji fonksiyonu ile global olarak en aza indirilmiştir. Bu çalışmada, ince taneli ürün tanıma için büyük ölçekli bir veri seti oluşturulmuş ve önerilen yöntemle bir temel oluşturulmuştur.

Diğer bir çalışmada [33], ürünler SIFT özelliklerine [56] dayalı bir olasılıksal çıkarım modeliyle olası kategorilerin kısa listesinin tahmin edilmesinden sonra, önceden eğitilmiş bir CNN modeli (VGG-f ağı [71]) ile yeniden sınıflandırılmıştır. Önerilen yöntem ile farklı veri kümeleri [6], [31] üzerinde önceki çalışmalar [6] ve [28]'e göre daha yüksek bir genel ortalama hassasiyet (mAP, mean Average Precision) değerine ulaşılmıştır. Bununla birlikte, perakende sektöründeki gerçek sorunları çözmek amacıyla bu tekniğin uygulanması için bu mAP değerlerinin hala önemli ölçüde iyileştirilmesi gerekmektedir [72]. Yine de önerilen model, derin tespit modellerinin eğitimi için yeterli veri toplanmasının mümkün olmadığı durumlarda kullanışlıdır. Ancak daha fazla eğitim verisi mevcutsa derin modeller daha iyi performans gösterir [73].

İnce taneli ürünleri tanımak için yeni bir yaklaşımın önerildiği çalışma [34]'te, raftaki potansiyel ürün dizilişlerine ve markalar arasındaki bütünleşik görsel hiyerarşiye göre bir güven seti oluşturulmuştur. Güven kümesinin sonuçlarının belirli bir güven seviyesinde doğru sınıfları içerdiği garanti edilmiştir. Çalışma [34]'te önerilen yöntem, en gelişmiş CNN modellerinden daha iyi performans

göstermiştir ancak, önemli ince taneli market ürünü veri kümeleri (örneğin Grocery Products [6]) üzerinde performansını kanıtlayamamıştır.

Bir diğer çalışma [35]'te, ortak ürün tespiti ve sınıflandırma modelinin göreve özgü eğitiminin performansını göstermek için farklı prosedürler önerilmiştir. İlk olarak, tüm ürünleri etiketlenmiş bir veri kümesi üzerinde bu ortak model eğitilmiştir. Daha sonra eğitim süreci, göreve özgü veri kümeleri üzerindeki eğitimin tamamlanması süresince değiştirilmiştir. Çalışmada kullanılan farklı prosedürlerle, eğitimi bir tespit ve sınıflandırma aşamasına ayırmanın, modelin geleneksel çok sınıflı bir detektör olarak eğitilmesine kıyasla performansında bir azalmaya neden olmadığı gösterilmiştir. Ortak ürün tespiti ve sınıflandırma modellerinin performansı, bir veri kümesinin tamamen etiketli olup olmamasına bağlıdır. Ayrıca gerçek market ürün tanıma problemlerinde raflardaki tüm ürünlerin etiketli olması, sürekli ürün eklenmesi ve yenilenen ürün ambalajları nedeniyle zorlu bir iştir.

Ürün tespitin ve sınıflandırmasının birlikte ele alındığı çalışma [36]'da raf görüntüsünde yer alan ürünlerin pozlarını tahmin etmek için Hough uzayında raf görseli ve ürün görsellerinin eşleşen SURF kilit noktaları üzerinden Hough oylama şeması oluşturulur. Sonrasında, Hough uzayında poz-sınıf histogramı kullanılarak ince taneli sınıflandırma işlemi gerçekleştirilir. Çalışma sonucunda önerilen yöntem kategorizasyon tahmininde yüksek başarı elde ederken, alt ürün kategorilerinin tahmininde aynı başarıyı gösterememiştir.

1.3.5 Çok Aşamalı Uçtan-uca Ürün Tanıma Çalışmaları

Ürün tanıma probleminin baştan sona ele alındığı üç aşamalı model [16], kısıtlamasız ürün tanıma, grafik tabanlı tutarlılık kontrolü ve ürün doğrulama aşamalarından oluşur. Kısıtlamasız ürün tanıma aşamasında özellik eşleştirme ve genelleştirilmiş Hough Dönüşümü [53] kullanılmıştır. Grafik tabanlı tutarlılık kontrolü kısmında, raflarda planlanan ürün pozisyonları (planogram) ile bir önceki aşamada elde edilen pozisyonlar arasındaki uyum kontrol edilmiştir. Son olarak ürün doğrulama için farklı yöntemler kullanılmış ve bunlar arasında en başarılı yöntem BRISK [11] olmuştur. Eğitim ve test veri setine ek olarak planogram bilgilerinin kullanılması, ürün konumlarının belirlenmesi ve kaçırılan ürünlerin yakalanması açısından daha iyi bir performans sağlamıştır. Ancak planogram bilgisinin elde edilemediği durumlarda beklenen performansa ulaşılması zordur.

Tablo 1.1 Literatürdeki market raflarında ürün tanıma çalışmalarının sınıflandırılması

Kaynak	Raf Tespiti	Ürün Tespiti	Ürün Sınıflandırma	Ürün Tespiti ve Sınıflandırmanın Ortak Olarak Gerçekleştirilmesi	Uçtan -uca	Çalışma Süresi Performansı
[31] (2007)	-	-	-	✓	✓	-
[45] (2007)	-	-	✓	-	-	-
[32] (2010)	-	-	-	✓	✓	✓
[6] (2014)	-	-	-	✓	✓	✓
[37] (2014)	✓	✓	✓	-	-	-
[17] (2015)	✓	✓	✓	-	-	-
[50] (2015)	-	-	***	-	-	✓
[36] (2016)	-	-	-	✓	✓	*
[39] (2016)	✓	✓	✓	-	-	-
[46] (2016)	-	-	✓	-	-	-
[15] (2017)	-	✓	✓	-	✓	*
[16] (2017)	-	✓	✓	-	✓	✓
[33] (2017)	-	-	-	✓	✓	✓
[47] (2018)	-	✓	✓	-	✓	*

Tablo 1.2 Literatürdeki market raflarında ürün tanıma çalışmalarının sınıflandırılması (devamı)

[5] (2019)	-	✓	-	-	-	✓
[34] (2019)	-	-	-	✓	✓	-
[38] (2020)	-	✓	-	-	-	✓
[40] (2020)	-	-	✓	-	-	-
[41] (2020)	-	✓	-	-	-	-
[42] (2022)	-	-	✓	-	-	✓
[43] (2022)	-	✓	-	-	-	✓
[48] (2022)	-	✓	**	-	✓	-
[49] (2022)	-	✓	**	-	✓	✓
[35] (2023)	-	-	-	✓	✓	-
[44] (2023)	-	-	✓	-	-	*
Yapılan tez çalışması	✓	✓	✓	-	✓	✓

✓ Satırda yer alan çalışmanın sütunda belirtilen ögeyi içerdiğini belirtir. * Satırda yer alan çalışmanın çalışma zamanı performans değerlendirmesiyle ilgili kısıtlı bilgi içerdiğini belirtir. ** Satırda yer alan çalışmada ürün sınıflandırması için metin tanıma kullanıldığını belirtir.

Başka çok aşamalı bir çalışma [15]'te farklı sayıda sınıfa ve farklı sayıda görüntüye sahip iki veri seti üzerinde alınan sonuçlar sunulmuştur. Önerilen yöntem ön aday seçimi, ince seçim ve işlem sonrası adımlarından oluşmaktadır. İnce seçim aşamasında önerilen iki yöntemden SURF özelliklerinin [10] kullanılması daha kolay veri setinde daha doğru sonuç verirken, bir derin sinir ağı (DNN, Deep Neural Network) olan AlexNet [64] ile elde edilen özellikler daha karmaşık veri setinde daha iyi bir performans göstermiştir. Önerilen yöntem, ilk aşamada her ürün için çok sayıda ayrı aday pencere oluşturulmasını gerektirdiği ve ikinci aşamada özelliklerin çıkarılması için gereken işlem süresinin artmasına neden olması nedeniyle oldukça karmaşıktır.

Başka bir çalışmada [47], ürün tespit adımı için özelleştirilmiş bir sadece bir kez bakarsın versiyon 2 (YOLOv2, You Only Look Once version 3) [29], tanıma için VGG-16 [74] ile öğrenilmiş tanımlayıcılar ve son olarak bazı iyileştirme stratejileri kullanılmıştır. Alınan sonuçlara göre çalışma [47]'de, [16] ve [28]'e kıyasla daha yüksek tanıma performansı kaydedilmiştir. Ancak YOLOv2'nin eğitilmesi için ekstra bir veri kümesine ihtiyaç duyulmaktadır. Ek olarak, market ürün tanıma veri kümeleri, yaygın olarak kullanılan nesne tanıma veri kümeleriyle karşılaştırıldığında daha fazla sayıda sınıfa ait daha az görüntüye sahiptir. Bu nedenle, market ürün tanıma veri kümeleri derin bir öğrenme modelini eğitmek için yeterli olmayabilir [72].

Diğer uçtan-uca çalışmalar [48] ve [49], ürünler üzerindeki metni tanıyarak ürün tanıma sorununu çözmeye çalışmıştır. Bu çalışmaların amacı, özellikle görme engelli kişilerin alışveriş deneyimini geliştirmektir. Çalışma [48]'in aşamaları ön işleme, ürün algılama ve ürün tanıma (metin algılama ve metin tanıma dahil) sırasıyla ele alınmıştır. Ürünleri tespit etmek için sadece bir kez bakarsın versiyon 5 (YOLOv5, You Only Look Once version 5) [75]; metin tespiti için ResNet50 [76] tabanlı bir FPN ve yeni bir işlem sonrası tekniği ve son olarak ürünlerin metin bilgilerini tanımak için seçici bağlam dikkatli metin tanıyıcı (SCATTER, Selective Context Attentional Text Recognizer) [77] kullanılmıştır. Önerilen yöntem mevcut tekniklerin etkinliğini arttırmaktadır.

Diğer bir çalışma [49], derin öğrenme tabanlı bir ürün tespiti aşaması ve görsel tabanlı özellik eşleştirmeye dayalı bir ürün sınıflandırma aşamasından oluşmaktadır. Ürün tespiti aşamasında Dense-Box stili mimarinin [78]

benimsendiği FPN kullanılmıştır. Ürün sınıflandırma aşamasındaysa, metin konumuna dayalı konumsal kodlama ile elde edilen özellik vektörleri Hungarian algoritması [79] ile eşleştirilmiştir. Bu çalışmada oluşturulan veri seti diğer veri setlerinden farklı olarak ürünle daha çok örtüşen, ürünün dört köşesinin koordinat bilgilerinden oluşur. Böylece, iyi hizalanmış, metinsel olarak geliştirilmiş büyük ölçekli veri seti ile metinden ürün tanıma çalışmaları için referans oluşturmuştur. Ambalajı çok benzer olan ürün çeşitlerini görsel özellikleri yerine metinden tanımak avantajlı olabilir. Ancak gerçek market raf görsellerinde ürünün konumuna veya fotoğrafın açısına bağlı olarak yazı görünür olmayabilir veya eksik görünebilir. Bu nedenle ürünlerin sadece metin tanıma ile tanınmaya çalışılması her durumda yeterli olmayabilir. Farklı görsel veya derin özelliklerle birlikte kullanımı performansı arttırabilir ama bu durumda da metinden ürünleri tanımak için önerilen bazı yöntemler hesaplama açısından maliyetli olması nedeniyle tercih edilmeyebilir.

1.4 Hipotez

Market raf görüntülerindeki ürünlerin tanınması için problemin farklı aşamalarının ayrı ayrı ele alınıp, her aşamaya uygun çözümler geliştirilmesinin sistemin performansını arttıracakı öngörülmektedir. Bunun için ürün tanıma süreci, öncelikle ürün konumlarının tespiti, daha sonra ürünlerin sınıflandırılması ve son olarak da iyileştirme adımıyla tanımlanır. Farklı aşamalar için geliştirilen uygun yöntemlerle oluşturulan uçtan-uca bir ürün tanıma sisteminin performansını kanıtlamak için literatürdeki tek aşamalı ürün tanıma sistemleri ile karşılaştırılmıştır.

Bu tez çalışmasında önerilen üründen bağımsız ürün tespit aşamasıyla, market raflarında ürün tanıma probleminin önemli zorluklarından biri olan sürekli olarak yeni ürünlerin eklenmesi ve ürün ambalaj tasarımlarının devamlı değişmesi nedeniyle oluşan sistemin yeniden eğitilmesi gereksiniminin ortadan kaldırılacağı varsayılmaktadır. Bunun için, ürün tespit modelinin eğitiminde test aşamasında kullanılmayan ürün çeşitlerinin yer aldığı farklı raf görüntüleri kullanılmıştır. Ayrıca, ürün tespiti aşamasında derin öğrenme modellerinin kullanımının ürün tanıma performansına etkisini göstermek için farklı bilgisayarla görme ve makine öğrenmesi yaklaşımları ile karşılaştırılmıştır.

Ürün sınıflandırma aşamasındaysa, ürün çeşitliliğinin fazla olmasından kaynaklanan farklı yöntemlerin farklı ürünlerde daha iyi sonuçlar göstermesi nedeniyle önerilen farklı özellik çıkarım yöntemlerinin bir arada kullanıldığı hibrit bir modelin sistem performansına artırıcı etkide bulunacağı öngörülmektedir. Bunun için, farklı özellik çıkarma yöntemlerinin tek başına kullanımının ve hibrit kullanımının ürün tanıma problemi üzerindeki performansı karşılaştırmalı olarak verilmiştir.

Çalışmanın iyileştirme aşamasında, ürün tespiti ve ürün sınıflandırma aşamalarından elde edilen her ürün için birden fazla lokalizasyon ve sınıflandırma sonuçlarının tek bir sonuca indirgenmesi için önerilen kümeleme algoritmasına eklenen komşuluk ilişkisi ile sistem performansının artırılacağı öngörülmektedir. Kümeleme algoritmasına komşuluk ilişkisi eklenilmesinin ürün tanımda etkisini göstermek için ilgili sonuçlar verilmiştir.

Otomatik bir ürün tanıma sisteminin gerçek hayat problemlerine cevap verebilmesi için performansı kadar çalışma süresinin de önemi vardır. Bu çalışmada, önerilen sistemin ve geliştirilen yöntemlerin sistem performansında yarattığı artırıcı etkinin yanında sistemin toplam çalışma süresinde önemli bir kayba neden olmayacağı öngörülmektedir. Bu nedenle, son olarak önerilen yöntemin çalışma zamanına etkisine de yer verilmiştir.

1.5 Tezin Genel Bakışı

Gerçekleştirilen bu tez çalışması aşağıda kısaca özetlenen beş bölümden oluşmaktadır:

Bölüm 1: Tezin giriş bölümü, tezin amacı, özgün değeri, literatür taraması, hipotez ve tezin genel bakışı kısımlarından oluşmaktadır. Tezin amacı kısmında market raflarında ürün tanıma problemine genel bir bakış sağlanmakta ve bu tez çalışmasının amacı ortaya koyulmaktadır. Literatür taraması kısmında market raflarında ürün tanıma problemiyle ilgili mevcut çalışmalar tanıtılmakta, sınıflandırılmakta ve konu ile ilgili literatürdeki boşluklara dikkat çekilmektedir. Hipotez başlığı altında ise bu tez çalışması kapsamında belirlenen hipotezler sunulmaktadır. Tezin genel bakışı kısmında ise tezin her bölümü ile ilgili özet bilgi verilmektedir.

Bölüm 2: Tezin ikinci bölümünde market raflarında ürün tanıma problemi, market raflarında ürün tanıma veri setleri ve market raflarında ürün tanıma için kullanılan bilgisayarla görme ve makine öğrenmesi yöntemleri başlıklarından oluşmaktadır. Market raflarında ürün tanıma problemi kısmında ürün tanımanın çözüm bulmaya çalıştığı farklı problemler hakkında bilgi verilmektedir. Market raflarında ürün tanıma veri setleri kısmında halka açık, kullanılabilir ürün tanıma veri setleri hakkında bilgi verilmiş, karşılaştırılmalı açıklamalara yer verilmektedir. Market raflarında ürün tanıma için kullanılan bilgisayarla görme ve makine öğrenmesi yöntemleri başlığı altında ürün tanıma probleminin farklı aşamaları için literatürde kullanılan yöntemler incelenerek yöntemlerin değerlendirilmesi yapılmaktadır.

Bölüm 3: Çok aşamalı uçtan-uca ürün tanıma sistemi bölümü, önerilen yöntem ve kullanılan veri setleri kısımlarından oluşmaktadır. Önerilen yöntem kısmında çok aşamalı ürün tanıma sisteminin her aşamasında kullanılan yöntemler tek tek açıklanmaktadır. Kullanılan veri setleri kısmında ise önerdiğimiz yöntemin eğitim ve test aşamalarında kullanılan veri setleri tanıtılmaktadır.

Bölüm 4: Deneysel sonuçlar kısmında önerilen yöntemlerin ilgili veri kümeleri üzerindeki deneysel sonuçları sunulmaktadır. Elde edilen sonuçlar, GP-20 veri setinde alınan sonuçlar, GP-181 veri setinde alınan sonuçlar, Grocery Products veri setinde alınan sonuçlar ve çalışma süresi performans değerlendirmesi başlıkları altında verilmektedir.

Bölüm 5: Son bölümde sunulan çalışma özetlenerek tezdeki bulguların ve katkıların genel değerlendirmesi yapılmış ve olası gelecek çalışmalar tartışılmaktadır.

MARKET RAFLARINDA ÜRÜN TANIMA

2.1 Market Raflarında Ürün Tanıma Problemi

Market raf görüntülerinde ürün tanıma problemi hem akademik hem de endüstriyel açıdan geniş araştırma çalışmaları yapılan önemli bir bilgisayarla görme ve makine öğrenmesi problemidir. Kullanıcıların ihtiyaçlarına ve kullanım alanlarına göre farklılık gösterse de raf görsellerindeki ürünleri tanıma sürecinde çözülmesi gereken iki temel problem vardır: (1) raftaki ürünün konumunun bulunması ve (2) ürünün sınıflandırılması. Sistemin çözüm bulması gereken sorunlara göre sadece ürünün sınıflandırılmasının, ya da sadece ürünün konumunun tespit edilmesinin, ya da ürünlerin hem konumunun hem de sınıf bilgisinin edinilmesinin gerekli olduğu durumlar mevcuttur. Şekil 2.1’de ürün sınıflandırması, ürün konumunun tespit edilmesi ve her ikisinin de birlikte gerçekleştirildiği durumlar farklı veri kümeleri üzerinde gösterilmiştir.



Şekil 2.1 Ürün tanıma problemi: Freiburg Groceries [46] veri setinden sınıflandırma örneği, SKU-110K [5] veri setinden ürün konumlarının tespiti örneği, Grocery Products [6] veri setinden ürün konumlarının tespiti ve sınıflandırılması örneği

Ürün sınıflandırması, bir görüntü veya video çerçevesindeki belirli ürünleri tanımlama işlemidir. Ürünü olabilecek markalar, ürün türleri veya farklı ürün kategorileri için önceden tanımlanmış sınıflarla eşleştirir. Örneğin “L’oreal” ve “Head&Shoulders” marka sınıflarıdır, “şampuan” ürün türü sınıfıdır ve “kişisel bakım ürünleri” ürün kategorisidir. Ürünün yalnızca rafta bulunup bulunmadığını

bilmenin önemli olduğu ancak rafta yer aldığı konum bilgisinin önemli olmadığı durumlarda ürün sınıflandırması esastır. WebMarket [45] ve Freiburg Groceries [46] veri kümeleri ürün sınıflandırması için özel olarak oluşturulmuş veri setleridir.

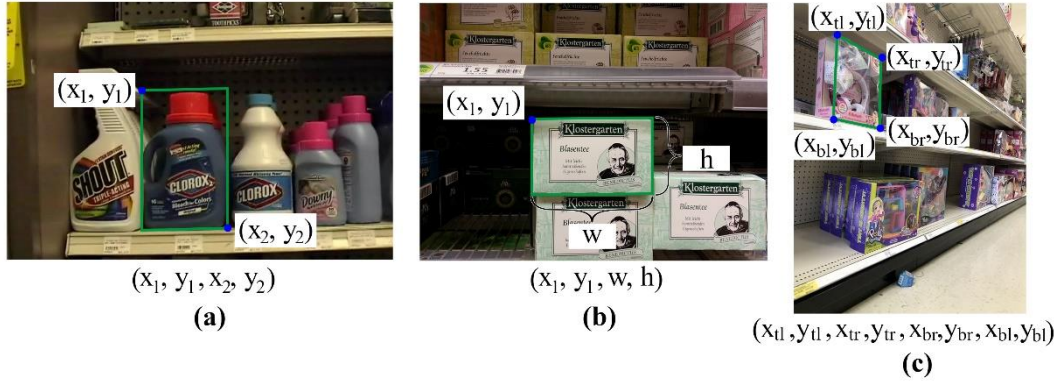
Ürün konumunun tespiti, raf görüntüsünde yer alan ürün konumunun kesin olarak belirlenmesini içerir. Bu, görüntü üzerinde yer alan her ürünün konumunu belirleyen sınırlayıcı kutu bilgilerinin tahmin edilmesi anlamına gelir. Ürün konumunun tespiti, raf görüntüsünde hangi ürünlerin mevcut olduğunun değil görselde nerede konumlandıklarını bilmenin gerekli olduğu senaryolar için önemli bir görevdir. SKU-110K [5] ve Unitail-Det [49] veri kümeleri, ürün konumu tespiti için özel olarak oluşturulmuş veri setleridir.

Ayrıca ürün sınıflandırma ve konum tespitinin birlikte ele alındığı problemler de vardır. Planogram eşleştirme ve stok takibi gibi ürün sınıfı ve ürün konum bilgisinin eş zamanlı olarak gerekli olduğu problemlerde kullanılır. Grozi-120 [31], Grocery Products [6], Grocery Dataset [17], CAPG-GP [80] ve Locount [81] gibi veri setleri ürünün hem sınıf ve hem de konum bilgisi ile etiketlenmiş veri setleridir.

Veri setlerinin konum bilgilerinin tutulma şekli veri setinden veri setine farklılık gösterebilmektedir. Şekil 2.2’de, farklı veri kümelerinden raf görsellerindeki ürün sınırlayıcı kutu bilgilerinin farklı gösterim formatları yer almaktadır. Örneğin, Grozi-120 [31], SKU-110K [5] ve Locount [81] veri kümeleri, ürünü çevreleyen dikdörtgen alanın sol üst (x_1, y_1) ve sağ alt (x_2, y_2) köşelerinin uzamsal koordinatlarını içeren açıklamalara sahiptir. Grocery Dataset [17] veri setinde ise aynı uzamsal koordinatlar farklı sırada (x_1, x_2, y_1, y_2) verilmiştir. Grocery Products [6] ve CAPG-GP [80] veri setlerindeki sınırlayıcı kutular ise sol üst (x_1, y_1) köşenin uzamsal koordinatı ile ürünün genişlik ve yükseklik değerleri ile ifade edilir. Hepsinden farklı olarak Unitail-Det’in [49] sınırlayıcı kutusu, sol-üst nokta (ptl, Point of Top Left), sağ-üst nokta (ptr, Point of Top Right), sağ alt nokta (pbr, Point of Bottom Right) ve sol alt nokta (pbl, Point of Bottom Left) olmak üzere sekiz serbestlik derecesine sahip $(x_{tl}, y_{tl}, x_{tr}, y_{tr}, x_{br}, y_{br}, x_{bl}, y_{bl})$ dört koordinat noktasından oluşur.

Market raflarında yer alan ürünleri sınıflandırmak ve konumunu tespit etmek müşterinin, üreticinin ya da tedarikçinin karşılaştığı birçok problemin çözümüne yardımcı olur. Yapılan çalışmalara bakıldığında, market raf görüntülerindeki

ürünlerin tanınmasının farklı kullanıcıların farklı ihtiyaçlarına cevap verebilecek daha spesifik problemler olarak ele alındığı görülmektedir. Market raf görüntülerinde ürünlerin tanınması ile çözülen problemler ve bunları ele alan çalışmalar Tablo 2.1’de listelenmektedir.



Şekil 2.2 Farklı veri kümelerindeki farklı sınırlayıcı kutu bilgisi gösterim şekilleri: (a) Grozi-120 [31] veri setinden sınırlayıcı kutu örneği, (b) Grocery Products [6] veri setinden sınırlayıcı kutu örneği, (c) Unitail-Det'in [49] veri setinden sınırlayıcı kutu örneği

Nesne alma, alışverişe yardımcı sistemler, ince-taneli sınıflandırma, planogram uyumluluğu, tek kare nesne algılama, stok takibi ve yoğun ürün algılama raf görüntülerindeki ürünlerin tanınması ile çözüm üretilen problemlerdir. Tablo 2.1’de listelenen çalışmalardan da görüldüğü üzere, bir çalışma birden fazla problemi ele alabilir ya da bu problemler birbirleriyle ilintili olabilir.

Nesne alma probleminde, sisteme bir ürünün sorgu görüntüsü verilir ve bu ürünün raf görüntüsünde bulunup bulunmadığı ya da bulunduğu konum bilgisi elde edilir [45]. Ürünün sorgu görüntüsü, ürünün raftan kesilmiş bir görüntüsü ya da ideal stüdyo ortamında çekilmiş görüntüsü ya da ürünün doğal ortamdaki görüntüsü olabilir.

Alışverişe yardımcı sistemler, sağladıkları bilgilerle müşterinin alışveriş deneyimini iyileştirmeyi amaçlamaktadır. Müşteriler alışveriş listesindeki ürünleri hızlı bir şekilde bulabilir, ürünlerin fiyatlarını karşılaştırabilir, aradığı ürünün rafta olup olmadığı ya da nerede olduğu hakkında bilgi edinebilir, ürün bulunmuyorsa mevcut alternatif ürünleri konum bilgileriyle birlikte tavsiye alabilir [36]. Ayrıca alışverişe yardımcı sistemler görme engelli kişilere kendi başına alışveriş yapma kolaylığı da sunmaktadır [32].

Tablo 2.3 Ürün tanımanın ele aldığı problemler

Ürün Tanımanın Ele Aldığı Problemler	Problemlerin Ele Alındığı Kaynaklar
Nesne Alma	[15] [45]
Alışverişe Yardımcı Sistemler	[6] [31] [32] [36] [44] [46] [47] [50]
İnce-taneli Sınıflandırma	[6] [33] [34] [47] [50] [80]
Planogram Uyumluluğu	[16] [17] [37] [39] [51] [52]
Tek Kare Nesne Algılama	[35] [42]
Stok Takibi	[81]
Yoğun Ürün Algılama	[5] [38] [41] [43]

İnce taneli sınıflandırma, geniş ürün kategorilerinin ötesine geçen ve belirli bir kategori içindeki ince farklar arasında ayırım yapmaya odaklanan bir sınıflandırma düzeyini ifade eder [6]. İnce-taneli sınıflandırma için bu yapıdaki veri setlerine ihtiyaç vardır. Örneğin, Grocery Products [6] veri seti 27 üst sınıfa bağlı 3.235 alt sınıf içerir. Veri setinde “peynir” sınıfı “süt ürünleri” ile süt ürünleri sınıfı ise “gıda” kategorisi ile bağlantılıdır. Bu süreç genellikle daha geniş ürün kategorilerinden daha spesifik sınıflara inilerek ilerler. Tanıma süreci sınıflandırmayı adım adım hassaslaştırdıkça daha hızlı ve daha güvenilir sonuçlar almayı sağlar.

Planogram uyumluluğu, perakende sektöründe perakendecinin veya markanın belirlediği planogramları ne ölçüde yansıttığını tanımlamak için yaygın olarak kullanılan bir terimdir. Planogram, ürünlerin mağaza raflarında nasıl düzenlenmesi ve sergilenmesi gerektiğini belirten görsel bir temsil veya diyagramdır [16]. Bu problemlerde veri setinin yanı sıra sistem girdi olarak planogram temsili ya da diyagramı, raf sayısı, raf başına düşen ürün sayısı veya ürün türü gibi planogram bilgilerine de ihtiyaç duyar.

Tek kare nesne algılama [82], geleneksel öğrenme modellerinden farklı olarak her bir kategori için sadece bir tane örnek görüntüye ihtiyaç duyar. Bu görüntülerden, her bir kategori için bir temsilci veya özellik vektörü oluşturur. Bu özellik vektörü

ile tanınacak öğeler arasındaki benzerliğe göre bir sınıflandırma yapılır. Tek kare nesne algılama, market ürünlerini tanıma gibi sınırlı veriye sahip olunan ve yeni kategorilere hızlıca adapte olma ihtiyacı duyulan durumlar için ele alınan bir yaklaşımdır.

Stok takibi, ürünün rafta bulunabilirliği ve ürün sayısının tutulması uygulamalarını içerir. Ürünün belirli bir raf görüntüsünde yer alıp almadığı, yer alıyorsa o üründen kaç tane bulunduğu tahmin edilir [81]. Böylece sistem, müşterilerin satın almak istediği anda ürünlerin ne ölçüde stokta bulunduğunu ve market raflarında bu ürünlerin satın alınıp alınamayacağını gösterir.

Yoğun ürün tespiti, raf görsellerinde çok sayıda birbirine yakın aralıklı ürünler ve sıklıkla benzer veya aynı görünen nesneler olduğunda kullanılır [5]. Çok sayıda ve benzer ürünlerin raflarda yoğun bir şekilde, bir arada yer aldığı ürün tanıma problemlerinde ele alınır.

Bilgisayarla görme ve makine öğrenmesi teknolojileri kullanılarak perakende sektöründe çözüm üretilen bu problemler sayesinde, operasyonel etkinlikler önemli ölçüde iyileşmekte, verimlilik ve müşteri memnuniyeti artmakta ve rekabet avantajı elde edilmektedir. Böylece, modern perakende işletmelerinin daha akıllı, hızlı ve müşteri odaklı bir hale gelmelerine katkıda bulunulur. Bu problemlere çözüm geliştirmek için önemli iki nokta vardır: (1) probleme uygun yöntemlerin ve (2) bu yöntemlerin test edileceği veri setlerinin belirlenmesi. Bu nedenle, ilerleyen bölümlerde market raflarında ürün tanıma veri setleri ve ürün tanıma için kullanılan yöntemler ele alınmıştır.

2.2 Market Raflarında Ürün Tanıma Veri Setleri

Market raflarında ürün tanımlama problemi, raf görüntüsündeki ürünlerin konum ve sınıf bilgilerinin elde edilmesini amaçlamaktadır. Bu amaçla geliştirilecek ürün tanıma algoritmalarını eğitmek ve test etmek için kullanılacak veri kümelerinin bulunması önemli bir gerekliliktir. Bu veri setleri, market raflarındaki ürünleri içeren görüntülerden ve bu ürünlerin etiketlenmiş bilgilerinden oluşur. Daha doğru ve sağlam modellerin geliştirilebilmesi için ürün tanıma veri setinin farklı aydınlatma koşullarında, farklı açılardan ve farklı uzaklıklardan elde edilmiş olması, değişik raf tasarımlarının ve ürün yerleşimlerinin kullanıldığı örneklerin yer

alması, yeterli ürün ve marka çeşitliliğinin sağlanması gerekir. Ayrıca, verilerin toplanmasında ve etiketlenmesinde uygun bir standardın belirlenmesi ve bu şekilde oluşturulan veri setlerinin kamuya açık hale getirilmesi hem yeni modellerin geliştirilmesini hem de diğer geliştirilen modellerle performans değerlendirme yapabilmeyi kolaylaştırır. Literatürde farklı ürün gruplarına, özelliklere, sınırlamalara ve karmaşıklıklara sahip çeşitli veri setleri mevcuttur. Raf görsellerinden ürün tanımaya yönelik halka açık veri kümeleri örnek görselleri Tablo 2.2’de, bu veri setlerinin ürün çeşitliliği, eğitim ve test setindeki görsel sayısı, açıklama dosyasının içeriği ve hangi ürünlere açıklamaların eklendiği hakkındaki bilgiler ise Tablo 2.3’te verilmektedir.

2.2.1 WebMarket

WebMarket veri seti [45] market ürünlerine ilişkin yayınlanan ilk veri setlerinden biridir. 102 farklı ürün çeşidinden 300 ürün şablon görüntüsü ve 3.153 raf görselinden oluşmaktadır. 2272×1704 veya 2592×1944 boyutundaki raf görüntüleri marketin 30 metre uzunluğundaki 18 parçalı 10 farklı rafından çekilen fotoğraflarla elde edilmiştir. Raf görüntüleri, ilk rafın en başından başlayarak son rafın son noktasına kadar ilerler. Birbirini takip eden görseller, her rafın tam olarak ele alındığını garantilemek için bir miktar üst üste binmektedir. Her raf görüntüsü, yaklaşık 1.5 metre yüksekliğinde ve 2 metre genişliğinde üç veya dört raf seviyesini kapsamaktadır. Raf görüntüleri elde edilirken çekim açısı veya uzaklık konusunda herhangi bir kısıtlama uygulanmamıştır. Raf görsellerinin aksine, şablon görsellerinin her biri düz bir arka plan önünde, farklı bakış açılarından veya uzaklıklardan her üründen 2 veya 3 örnek içerecek şekilde oluşturulmuştur. Toplamda 300 şablon görüntüsü düz bir arka plan önünde kaydedilmiştir. WebMarket veri setinden şablon ve raf görsellerine ait örnekler Tablo 2.2’de verilmiştir. Veri setinin açıklama dosyası sadece her ürünün yer aldığı raf görüntüsü bilgisini içermektedir, ürünlerin konum bilgisi bulunmamaktadır.

2.2.2 Grozi-120

Grozi-120 veri seti [31] in vitro (eğitim için şablon görüntüleri) ve in situ (test için raf görüntüleri) görüntülerden oluşur. Hem nesnelerin sınıflandırması hem de konumlarının tespiti için oluşturulmuş ilk veri setidir. Web’den derlenmiş, in vitro olarak adlandırılan, 120 farklı ürün sınıfına ait toplam 676 adet ürün görseli

içermektedir. Her ürün sınıfını temsil eden minimum 2 ve maksimum 14 görsel vardır. Test görüntüleri de, 29 farklı market rafının 30 dakikalık bir videosundan her 5 karede bir çekilen 11.194 raf görüntüsünden oluşmaktadır. Raf görüntülerinde her ürün sınıfından en az 14, en fazla 814 örnek bulunur. Ürün şablon görselleri, çeşitli satıcılardan veya fotoğraf galerilerinden toplanan farklı boyutlarda, pozlarda ve aydınlatma koşullarındaki görüntülerden oluşmaktadır. Raf görselleri ise market koşullarındaki aydınlatma, ölçek, yansıma, poz, renk ve rotasyondaki çeşitliliklerle elde edilmiştir. Şablon ve rafların örnek görselleri Tablo 2.2'de verilmiştir. Raf görsellerinin temel gerçek sınırlayıcı kutu bilgileri, sol üst ve sağ alt köşelerinin uzamsal koordinatlarını içerir.

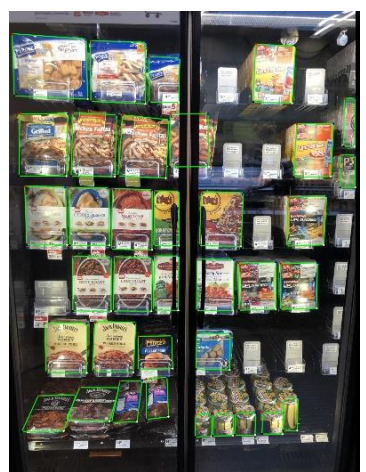
2.2.3 Grocery Products

Grocery Products [6] veri seti ince taneli yapısından dolayı diğer veri setlerinden farklıdır. Her ürüne ilişkin hiyerarşik kategorik bilgileri içerir. Bu bilgiye göre her ürünün bir veya daha fazla üst kategorisi bulunmaktadır. Grocery Products veri kümesinde [6] 8.350 alt kategoriye sahip 80 geniş ürün kategorisi bulunmaktadır. Ancak bu 80 ürün kategorisinden yalnızca 27'si etiketlenip açıklama dosyaları oluşturulmuştur. Dolayısıyla bu veri tabanı ile çalışan tüm çalışmalarda yalnızca bu 27 kategori dikkate alınmaktadır. Grocery Products [6] eğitim seti, web'den toplanan 27 üst ürün kategorisi altında 3.235 ayrıntılı ürün şablonu görüntüsünden oluşmaktadır. Her ürün şablonu, stüdyo benzeri koşullarda beyaz arka planda çekilmiş tek bir görselle temsil edilmektedir. Her ince taneli ürün kategorisinde ortalama 112 farklı perakende ürün içeren 25 ile 415 arasında eğitim fotoğrafı bulunur. Test seti, çeşitli aydınlatma koşulları, görüntüleme perspektifleri ve yakınlaştırma seviyeleri ile gerçek marketlerden cep telefonu ile çekilen 680 raf görüntüsünden oluşmaktadır. Her test görselindeki ürün sayısı 6 ile 30 arasında değişmektedir. Tablo 2.2'de Grocery Products veri setine ait eğitim ve test görsellerinin örnekleri gösterilmektedir. Her test görüntüsünün açıklama dosyası temel gerçek sınırlayıcı kutu ve kategori bilgilerinden oluşur. Ancak temel gerçek sınırlayıcı kutular tek bir ürünü değil aynı ürünün birden çok örneğini sınırlamak için kullanılır.

Tablo 2.1 Market ürünleri tanıma veri setlerine ait örnek görüntüler

Veri Seti	Eğitim Seti Örneği	Test Seti Örneği
WebMarket [45]		
Grozi-120 [31]		
Grocery Products [6]		
Grocery Dataset [17]		
Freiburg Groceries [46]		

Tablo 2.2 Market ürünleri tanıma veri setlerine ait örnek görüntüler (devamı)

Freiburg Groceries [46]		
CAPG-GP [80]		
SKU-110K [5]		
Locount [81]		
Unitail-Det [49]		

Tablo 2.3 Market ürünleri tanıma veri setleri

Veri Seti	Ürün Çeşidi	Eğitim Seti	Test Seti	Açıklama Dosyası	Açıklamalı Ürünler
WebMarket [45]	102	300 şablon görüntü (ürün başına 2-3 görsel)	3.153 raf görüntüsü	ürün sınıfı	seçili ürünler
Grozi-120 [31]	120	676 şablon görüntü (ürün başına 2-14 görsel)	11.194 raf görüntüsü	ürün sınıfı + sınırlayıcı kutu	seçili ürünler
Grocery Products [6]	27 üst-sınıf	3.235 şablon görüntü (ürün başına 25-415 görsel)	680 raf görüntüsü	ürün sınıfı + sınırlayıcı kutu*	seçili ürünler
	3235 alt-sınıf	3.235 şablon görüntü (her ürün için 1 görsel)	680 raf görüntüsü	ürün sınıfı + sınırlayıcı kutu*	seçili ürünler
Grocery Dataset [17]	10	3.701 şablon görüntü (ürün başına 274-598)	354 raf görüntüsü	ürün sınıfı + sınırlayıcı kutu	tüm ürünler
Freiburg Groceries [46]	25	4.947 ürün görüntüsü	74 raf görüntüsü	ürün sınıfı	seçili ürünler
CAPG-GP [80]	102	177 ürün görüntüsü (ürün başına 1-4 görsel)	234 raf görüntüsü	ürün sınıfı + sınırlayıcı kutu	seçili ürünler

Tablo 2.4 Market ürünleri tanıma veri setleri (devamı)

SKU-110K [5]	1	eğitim için 8.233 + doğrulama için 588 raf görüntüsü	2.941 raf görüntüsü	sınırlayıcı kutu	tüm ürünler
Locount [81]	140	34.022 raf görüntüsü	16.372 raf görüntüsü	ürün sınıfı + sınırlayıcı kutu* + ürün sayısı	tüm ürünler
Unitail-Det [49]	1	eğitim için 8.216 + doğrulama için 588 raf görüntüsü	başlangıç etki alanı için 2.940 + çapraz etki alanı için 500 raf görüntüsü	sınırlayıcı kutu	tüm ürünler

* Tek bir sınırlayıcı kutu alanının tek bir ürünü değil, birden çok ürünü kapsadığını belirtir.

2.2.4 Grocery Dataset

Grocery Dataset veri kümesi [17], sadece sigara paketlerini içeren daha spesifik bir veri kümesidir. Veri seti, 10 farklı kategorideki ürün görsellerinden ve açıklama dosyasına sahip 354 raf görüntüsünden oluşmaktadır. Çeşitli marketlerden alınan her sınıfa ait 5 ürün paketi, beyaz zemin üzerinde 4 farklı kamera kullanılarak fotoğraflanmıştır. Veri setinde çeşitlilik sağlamak amacıyla 40 farklı marketten 4 farklı kamera ile raf görüntüleri alınmıştır. Her görselde 2 ila 7 adet raf ve 2 ila 137 adet ürün bulunmaktadır. Her seferinde bir raf aşağı kaydırılarak rafların fotoğrafı çekilmiştir. Çekilen 354 raf görselinde toplamda 13.000 civarında ürün bulunmaktadır. Ürün ve raflara ait örnek görseller Tablo 2.2'de verilmiştir. Veri seti ile birlikte verilen raf görsellerinin temel gerçek sınırlayıcı kutu bilgileri ürünün sol üst köşesinin uzamsal koordinatlarından ve ürünün genişlik ve yükseklik değerlerinden oluşmaktadır.

2.2.5 Freiburg Groceries

Freiburg Groceries [46] veri setinin eğitim görüntüleri (D1) sadece marketlerde değil farklı ortamlarda kullanılarak dört ayrı kamerayla çekilmiştir. Eğitim seti,

256x256 boyutunda, 25 ürün kategorisinde, sınıf başına 97 ile 370 arasında örnekten, toplam 4.947 görüntüden oluşur. Freiburg Groceries test seti (D2), laboratuvar ortamındaki darmadağın sahneler, dağınık arka planlar ve farklı aydınlatma koşullarında elde edilmiştir. Elde edilen veri seti, 1920x1080 çözünürlüklü her bir raf görüntüsüne karşılık gelen sahnenin bir nokta bulutunu, bir RGB görüntüsünü ve bir derinlik görüntüsünü içerir. Her test görüntüsü birden fazla sınıfa ait çeşitli ürünler içerir. Eğitim ve test görsellerine ait örnekler Tablo 2.2'de verilmiştir. Veri setinin açıklama dosyası her test görselinin sadece etiketli ürün sınıfı bilgilerinden oluşmaktadır, ürünlerin sınırlayıcı kutu bilgileri mevcut değildir.

2.2.6 CAPG-GP

CAPG-GP [80] ince taneli bir ürün tanıma veri setidir. Eğitim seti, e-ticaret sitelerinden derlenen 177 adet stok tutma birimi (SKU, Stock Keeping Unit) görüntüsünden oluşmaktadır. Eğitim görselleri üç farklı ürün türünden oluşmaktadır: tüp ambalajlı ürünlerden 18 sınıf, poşet ambalajlı ürünlerden 15 sınıf ve kutu ambalajlı ürünlerden 69 sınıf. Toplamda 5 farklı markanın en az 1, en fazla 4 örnekle temsil edildiği 102 ürün sınıfından toplam 177 görüntüden oluşmaktadır. Test seti ise, 2 marketten cep telefonu kameralarıyla fotoğraflanan 234 raf görüntüsünden oluşmaktadır. Her raf görselinde ortalama 20 adet ürün bulunmaktadır. Tablo 2.2'de eğitim ve test görüntülerinin örnekleri gösterilmektedir. Raf görsellerinin açıklama dosyasında yalnızca 102 ürünün sınırlayıcı kutu bilgisi yer almaktadır.

2.2.7 SKU-110K

SKU-110K [5] hem eğitim hem de test setinin raf görüntülerinden oluştuğu oldukça büyük ölçekli bir veri setidir. Ayrıca SKU-110K [5], raflardaki ürün sayısı ve çeşitliliği ile yoğun yapısı nedeniyle farklılık göstermektedir. Raf görsellerindeki ürün alanının ortalaması yaklaşık %0,27'dir. Görüntüler Amerika Birleşik Devletleri, Avrupa ve Doğu Asya'daki binlerce süpermarkette cep telefonlarıyla çekilmiştir. Raf görselleri farklı ölçeklere, görüş açlarına, aydınlatma koşullarına, gürültü seviyesine ve farklı öngörülemez koşullara sahiptir. Çünkü çekim sırasında fotoğraf kalitesi ve görünüm ayarlarıyla ilgili herhangi bir düzenleme yoktur. Veri seti, %70'i (8.233 görüntü – 1.210.431 sınırlayıcı kutu) eğitim, %5'i (588 görüntü – 90.968 sınırlayıcı kutu) doğrulama ve %25'i (2.941 görüntü – 432.312 sınırlayıcı

kutu) test olmak üzere toplam 11.762 görüntüden oluşur. Veri kümesinin bölünmesi, aynı markete ait aynı raf teşhirlerinin aynı alt kümelerde olmasını engellemek amacıyla rastgele seçim ile gerçekleştirilmiştir. Veri setinde rafta yer alan her ürün, sınırlayıcı kutu ile tek sınıf olarak etiketlenmiş ve ürün çeşitleri belirtilmemiştir. SKU-110K veri setinin [5] örnek görselleri Tablo 2.2'de verilmiştir.

2.2.8 Locount

Locount [81] 28 farklı marketten, farklı görüş açıları ve aydınlatma koşullarında çekilmiş 50.394 raf görüntüsünden oluşur. Eğitim seti 1.437.166 örnekle 34.022 fotoğraftan oluşurken, test seti 468.151 örnekle 16.372 görselden oluşmaktadır. Locount veri setinin örnek görselleri Tablo 2.2'de verilmiştir. Bebek eşyaları, içecekler, gıda maddeleri, günlük kimyasallar, giyim, elektrikli ev aletleri, saklama gereçleri, mutfak gereçleri ile kırtasiye ve spor malzemeleri olmak üzere 9 geniş farklı nesne kategorisinde ince taneli bir yapıya sahiptir. Bu 9 geniş kategorinin toplam 140 alt kategorisi bulunmaktadır. Eğitim ve test setinin açıklama dosyası, raf görsellerindeki her nesne sınıfına ait sınırlayıcı kutulardan ve sınırlayıcı kutunun içindeki örnek sayısı bilgisinden oluşur. Veri kümesi farklı ölçeklerde çeşitli nesneler içerir ve nesne ölçeklerinin dağılımı, sınırlayıcı kutulardaki piksel sayısına bağlı olarak küçük ölçek, orta ölçek ve büyük ölçek olmak üzere üç alt kümeye ayrılmaktadır. Veri setinin %52,5'i orta ölçekli ürünlerden oluşurken, %29,5'i büyük ölçekli, %18'i ise küçük ölçekli ürünlerden oluşmaktadır. Ayrıca, açıklamalı sınırlayıcı kutuların içerdiği örnek sayıları çoğunlukla tek, bazılarında 2 ila 10 örnek arasında ve az sayıda varyantta da 10'dan fazla örnek içermektedir.

2.2.9 Unitail-Det

Unitail-Det [49], ürün konumlarının tespiti için uzmanlaşmış birleşik perakende veri kümelerinden biridir. İki kaynaktan oluşturulmuştur: başlangıç etki alanı ve çapraz etki alanı. Başlangıç etki alanında, eğitim ve test fotoğrafları aynı mağazalarda, aynı sensörler aracılığıyla karşılaştırılabilir açılardan elde edilen görüntülerden oluşur. Başlangıç etki alanını oluşturmak için SKU-110k veri setinden [5] 11.744 fotoğraf seçilmiştir. Başlangıç etki alanı eğitim setinde 1.215.013 örnekle açıklamalı 8.216 görüntü bulunurken, doğrulama kümesinde 92.128 örnekle 588 görüntü yer almaktadır. Başlangıç etki alanı test seti, 432.896

açıklamalı 2.940 görüntüden oluşuyor. Çapraz etki alanında ise, çeşitli marketlerden kullanılmayan kategorilerden, çeşitli sensörler ve kamera perspektifleri kullanılarak 3024x4032 boyutunda 500 fotoğraf çekilmiştir. Görüntülerde şarküteri ürünlerinden ev eşyalarına kadar geniş bir ürün yelpazesinde toplam 1.454 ürün sınıfı yer almaktadır. Tablo 2.2’de, Unitail-Det’in eğitim ve test görüntülerinin örnekleri gösterilmektedir. Bu veri seti, ürünler için sınırlayıcı kutuların ön yüzünü temsil eden dörtgen sınırlayıcı kutu bilgi dosyası ile farklılık yaratmaktadır. Sınırlayıcı kutular sekiz serbestlik derecesine sahip $(x_{tl}, y_{tl}, x_{tr}, y_{tr}, x_{br}, y_{br}, x_{bl}, y_{bl})$ dört koordinattan oluşmaktadır.

2.3 Market Raflarında Ürün Tanıma için Kullanılan Bilgisayarla Görme ve Makine Öğrenmesi Yöntemleri

Market raflarında ürün tanıma probleminde bilgisayarla görme ve makine öğrenmesi yöntemlerinin kullanımı birçok avantaj sağlar. Makine öğrenmesi yöntemleri büyük veri setleri üzerinden öğrenerek ve verileri hızlı bir şekilde işleyerek ürünleri tanıyabilir. Veri seti genişledikçe ve model daha fazla öğrendikçe, sistemlerin ürün tanıma yetenekleri de artar. Ek olarak, iyi performans gösteren bir tanıma sistemi, insandan kaynaklı hataları azaltabilir ve daha tutarlı bir performans sunabilir. Böylece, iş süreçlerinin hızlanmasına ve verimliliğin artmasına katkı sağlar. Ayrıca, verinin elde edilmesi ve işlenmesi için gerekli donanım, yazılım yapılarının maliyetinin azaltılması ve verinin daha erişilebilir olması bu yöntemlerin uygulanmasını kolaylaştırmıştır. Bu avantajlar, perakende sektöründe bilgisayarla görme ve makine öğrenimi tabanlı sistemlerin benimsenmesine yönelik eğilimi daha da artırmıştır. Uçtan-uca bir ürün tanıma sisteminin çözüm bulması gereken problemlerden ürün konumlarının tespiti, ürünlerin sınıflandırılması ve sonuçların iyileştirilmesi gibi aşamalarda kullanılan bilgisayarla görme ve makine öğrenmesi yöntemleri bu bölümde ele alınmıştır.

2.3.1 Ürün Tespiti için Kullanılan Yöntemler

Market raflarındaki ürünleri tanımak için hangi ürün çeşidi olduğuna bakılmaksızın sadece ürün konumunun tespit edilmesini amaçlayan ya da ürün konumlarının ürün sınıfı bilgisiyle tespit edilmesini amaçlayan yaklaşımlar mevcuttur. Tablo 2.4’te ürün tespiti için kullanılan algoritmalar ve kullanıldığı çalışmalar listelenmiştir. Kullanılan algoritmalar arasında Kademeli Nesne Detektörü [55], Faster R-CNN

[28], RetinaNet [58], CenterNet [60], Cascade R-CNN [61], YOLOv2 [29], YOLOv5 [75] ve DenseBox [78] algoritmaları yer almaktadır.

Tablo 2.1 Ürün tespiti için kullanılan yöntemler

Ürün Tespiti için Kullanılan Yöntemler	Yöntemlerin Kullanıldığı Kaynaklar
Kademeli Nesne Detektörü [55]	[37] [39]
Faster R-CNN [28]	[35]
RetinaNet [58]	[5] [41]
CenterNet [60]	[38]
Cascade R-CNN [61]	[43] [81]
YOLOv2 [29]	[47]
YOLOv5 [75]	[48]
DenseBox [78]	[49]

Kademeli Nesne Detektörü [55], nesne tespiti için birden fazla aşamalı bir yaklaşımı benimser. Her aşama, önceki aşamadan daha fazla bilgi ve hassasiyet kazanmaya odaklanarak nesneleri daha iyi tespit etmeyi amaçlar. İlk aşamada, daha hızlı ve düşük maliyetli tespitleri gerçekleştirir. Sonraki aşamalarda, önceki aşamada yapılan hataların bir sonraki aşamada düzeltilmesi ile performansı arttırarak ilerler. Kademeli Nesne Detektörünün [55], özellikle zorlu nesnelerin tespiti veya daha yüksek hassasiyet gerektiren market ürünlerini tanıma gibi görevlerde [37] [39] kullanımı etkilidir. Fakat kademeli yapısı nedeniyle ilerleyen kademelerde sistem performansını arttırsa da çalışma süresinin artmasına neden olur.

Faster R-CNN [28], R-CNN'nin [83] geliştirilmiş bir versiyonudur. R-CNN'de [83] sunulan iki aşamalı model, Faster R-CNN'de [28] tek bir aşamada birleştirilerek konum tespiti ve sınıflandırması aynı anda yapılır. Bölge önerileri oluşturmak için R-CNN'de kullanılan seçici arama algoritması yerine Faster R-CNN'de [28] bölge

teklif ağı (RPN, Region Proposal Network) kullanılarak sistemin hesaplama maliyeti azaltılmış, daha hızlı ve daha etkili bir nesne tespit modeli sunulmuştur.

RetinaNet [58] yoğun nesne algılamada sınıf dengesizliği sorununu çözmek için hem nesnelerin konum tespitini hem de sınıflandırmasını tek bir ağda birleştiren bir algoritmadır. Odak kaybı adı verilen özel bir kayıp fonksiyonu kullanarak sınıflandırılması zor örneklerle daha yüksek ağırlıklar atayıp, nadir sınıflardaki nesnelerin daha iyi öğrenilmesini sağlar [58].

CenterNet [60] tahmin edilen merkez noktaları etrafında nesneleri modelleyerek, sınırlayıcı kutuları bu merkezlerden gerileyerek elde eder. Nesne merkezini tahmin etme, boyut ve sınıfı belirleme gibi görevleri tek aşamada birleştirir. İki aşamalı detektörlere kıyasla daha basit bir mimariyle daha rekabetçi bir doğruluk elde eder.

Cascade R-CNN [61], kademeli yapısı ile kademeli nesne detektörüne [55] benzer ve R-CNN yapısının bir varyantını temsil eder. Cascade R-CNN, iki aşamalı R-CNN mimarisinin kademeli yapıyla genişletilmesi ile özellik çıkarımı için bir omurga ağından, bölge teklifleri oluşturmak için bir RPN'den ve nesne tespitinin birden fazla aşamasından oluşur. Her aşamada model, önceki aşamadaki zorlu örneklerle veya yanlış pozitiflere odaklanacak şekilde ince ayar yapılır. Bu kademeli aşamalar, modelin zorlu nesneleri daha iyi tespit etmesini ve sınıflandırmasını sağlar [61].

YOLOv2 [29], YOLO mimarisini birden fazla sınıfta nesne algılamayı yönetecek şekilde genişletir. Bir görüntüyü tek bir geçişte analiz ederek nesneleri tespit eder. Doğruluğu artırmak için bağlantı kutuları ve hiyerarşik sınıflandırma sunar. YOLOv5 [75], YOLOv2'nin hiyerarşik yapısından uzaklaşır ve daha geleneksel bir CNN mimarisine yönelir. YOLOv5 hız ve hassasiyet açısından iyileştirmeler sunar. Ayrıca, YOLOv5 modelini özelleştirmek ve yeni sınıflar eklemek daha kolaydır [75].

DenseBox [78], nesne yoğunluğuna sahip durumlarda nesne tespiti için kullanılan bir algoritmadır. Yoğun kutu oluşturma stratejisi ile nesneleri sınıflandırır ve konumlarını tahmin eder. Geleneksel yöntemlerden farklı olarak, yoğun bölge üretimiyle daha fazla aday bölge elde edilir. Her bir yoğun öneri bölgesi bir CNN kullanılarak özellik haritasına dönüştürülür. Bu durum yoğun sahneleri tespit etme performansını artırırken, çalışma süresi açısından dezavantaj yaratır [78].

Bu yöntemler, her biri kendine özgü özelliklere, avantajlara ve dezavantajlara sahip nesne tespitine yönelik çeşitli yaklaşımları temsil eder. Ürün tespiti için belirli bir yöntemin seçimi, görevin gereksinimleri, veri kümesinin özellikleri ve hesaplama kısıtlamaları gibi çeşitli faktörlere bağlıdır.

2.3.2 Özellik Çıkarma Yöntemleri

Temel bir ürün tanıma algoritması, raf görüntüsündeki potansiyel ürün konumlarını bulur ve bu potansiyel alanı ürün şablonu görselleriyle eşleştirerek bu konumdaki ürünün sınıfını belirler. Bu süreçler birlikte veya ayrı aşamalarda ele alınır. Her iki durumda da market ürünlerinin konumunu tespit etmek ve tanımak için görsellerin hem görsel hem de metinsel özelliklerinden faydalanılır. Tablo 2.5'te listelendiği gibi, bu özellikler, geleneksel özellik çıkarma yöntemleri veya derin öğrenme tabanlı mimarilerle çıkarılmıştır. Daha iyi tanıma performansı elde etmek için, otomatik ürün tanıma sistemleri, farklı ürün tanıma kategorileri için birkaç özellik çıkarma yöntemini ayrı ayrı veya birlikte kullanabilir. Niteliklerin seçimi ve bunların kombinasyonu, ürün tanıma görevinin özel gereksinimlerine ve zorluklarına bağlıdır. Yani sistemler, verilerin kullanılabilirliği, ürünlerin doğası, görüntülerin ve metnin kalitesi, ürün tanıma hızı ve istenen tanıma doğruluğu düzeyi gibi faktörlere bağlı olarak farklı özellik kombinasyonlarını kullanabilir.

Tablo 2.1 Özellik çıkarma yöntemleri

Özellik Çıkarma Yöntemleri	Yöntemlerin Kullanıldığı Kaynaklar
Geleneksel özellik çıkarma yöntemleri	[6] [15] [16] [17] [31] [32] [33] [34] [36] [37] [39] [45] [50] [80]
Derin öğrenme tabanlı özellik çıkarma yöntemleri	[15] [40] [42] [46] [47] [48] [49] [81]

2.3.2.1 Geleneksel Özellik Çıkarma Yöntemleri

Geleneksel özellik çıkarma yöntemleri, renk özellikleri, yerel özellikler ile şekil ve çizgi özellikleri gibi öznitelikler üzerinden ürünleri kapsamlı bir şekilde anlamayı sağlar. Market raflarındaki ürünlerin tanınması için kullanılan geleneksel özellik çıkarma yöntemleri ve bu yöntemlerin kullanıldığı çalışmalar Tablo 2.6'da verilmiştir.

Görüntüde renk özellikleri, farklı renk uzayları ile temsil edilir. Lab renk uzayı ve HSV renk uzayı market raflarında ürün tanıma probleminde tercih edilen renk uzaylarındandır. Lab renk uzayı, renkleri temsil ettiği üç değerle insan gözünün algılamasına yakın bir modelle ifade eder [84]. Bu üç değerden "L" değeri, renk parlaklığını temsil eder ve 0 ile 100 arasında değişir. 0, tamamen siyahı; 100, tamamen beyazı temsil eder ve aradaki değerlerde grinin geçiş tonlarını oluşturur. "a" değeri yeşil-kırmızı eksen üzerinde, renk tonunun yeşilden kırmızıya olan konumu temsil eder. Değer aralığı -128 ile +127 arasındadır. Negatif değerler yeşil tonlarına işaret ederken pozitif değerler ise kırmızı tonlarına işaret eder. "b" değeri ise mavi-sarı eksen üzerinde, renk tonunun maviden sarıya olan konumu temsil eder. "b" değer aralığı da -128 ile +127 arasında değişir ve negatif değerler mavi tonlarına işaret ederken pozitif değerler de sarı tonlarına işaret eder.

HSV renk uzayı renkleri temsil etmek için renk özü, doygunluk ve değer bileşenlerini kullanır [84]. Renk değeri 0'ın kırmızıyı, 120'nin yeşili, 240'ın da maviyi temsil edeceği şekilde 0 ile 360 derece arasında bir değerle ifade edilir. Doymunluk değeri, 0 rengin en soluk halini, 100 ise rengin en canlı halini temsil edecek şekilde 0 ile 100 arasında değer alır. Değer ifadesi, Lab renk uzayındaki "L" değeri ile aynı şekilde rengin parlaklığını ifade eder ve aynı değerlerle temsil edilir. Market raflarında ürün tanıma problemi için ürünün şablon görüntüsü ile raf görüntüsünden elde edilen bölge önerileri arasındaki benzerlik, görüntülerin renk uzayları ile oluşturulan histogramların kesişimi ya da hesaplanan uzaklık metrikleri üzerinden hesaplanır.

Görüntülerin yerel özellikleri, görüntünün belirli bölgelerindeki ayırt edici desenleri veya yapıları ifade eden anahtar nokta tanımlayıcılarıdır. Ürün tanıma probleminin çözümünde yaygın olarak kullanılan yerel özellik çıkartıcıları arasında SIFT [56], SURF [10], BRISK [11], ORB [14], AKAZE [85] bulunur. SIFT [56] görüntülerdeki özellikleri ölçek uzayında farklı ölçek seviyelerinde tespit eder ve ardından bu özellikleri çıkartarak görüntüyü temsil eden benzersiz bir özellik vektörü oluşturur. Bu özellik vektörleri, ürün tanımda ürün şablonu ve raftaki ürün arasındaki benzerlik karşılaştırmaları için kullanılır.

SIFT [56] sağladığı ölçek değişimlerine ve rotasyona karşı değişmezlik özelliği ile farklı durumlarla başa çıkabilir. Nesnelerin farklı ölçeklerde görüldüğü durumları ele alarak özelliklerin ölçek değişimlerine karşı duyarlılığını azaltır. Nesnelerin

farklı yönler'e döndüğü durumları ele alarak da özelliklerin rotasyona karşı duyarlılığını azaltır.

SURF [10], SIFT'e benzer ancak hesaplama açısından daha verimli bir şekilde tasarlanmıştır. SURF'te kullanılan integral görüntü [55], hessian matrisi gibi parametre sayısını ve bellek kullanımını azaltan ve işlemleri hızlandıran avantajlara sahiptir.

Tablo 2.2 Geleneksel özellik çıkarma yöntemleri

Geleneksel Özellik Çeşitleri	Geleneksel Özellik Çıkarma Yöntemleri	Yöntemlerin Kullanıldığı Kaynaklar
Renk Özellikleri	Lab	[31] [32]
	HSV	[17] [37]
Yerel Özellikler	SIFT [56]	[6] [16] [17] [31] [33] [34] [39] [45] [80]
	SURF [10]	[15] [16] [32] [36] [50] [80]
	BRISK [11]	[16] [80]
	ORB [14]	[16]
	AKAZE [85]	[16]
Şekil ve Çizgi Özellikleri	Harris-Affine [63]	[45]
	Boosted-Haar [55]	[31]
	HoG [54]	[17] [37]
	Hough Dönüşümü [53]	[16] [36]

BRISK [11], adaptif çekirdek boyutları kullanarak farklı ölçek seviyelerinde özellikleri temsil eden skalar anahtar noktaları seçer. BRISK özellikleri [11] ölçek, rotasyon ve sınırlı afin değişikliklere göre değişmezdir. Rotasyon değişmezliğine

ulařmak için her bir özelliğın karakteristik yönünü tanımlar. Kullanılan ikili açıklamalarla da özelliklerin daha etkili bir şekilde eşleştirilmesini sağlar.

ORB [14], hızlı ve etkili bir özellik tespit yöntemi olan FAST'ı kullanır. FAST, bir pikselin etrafındaki 16 pikseli kontrol eder ve bu piksellerin yoğunluğunu karşılaştırarak belirli bir pikselin kenar veya köşe olup olmadığını belirler. Harris Corner skorunu kullanarak bu kenar ve köşeleri filtreler. Sonrasında BRIEF tanımlayıcıları ile elde edilen özelliklerle birlikte ikili vektörle temsil edilen ORB tanımlayıcısı [14] oluşturulur. ORB [14], SIFT [56] ve SURF'e [10] alternatif olarak kullanılan hızlı ve verimli bir yöntemdir.

AKAZE [85], doğrusal olmayan ölçek uzayı oluşturabilme yeteneğine dayanır. Hessian matrisi ve determinantı kullanılarak hassas bir özellik tespit yöntemini benimser. Binominal filtreleri ve kurtosis maksimizasyonu yöntemi kullanılarak özellikler belirlenir. Her bir özellik, çevresindeki piksellerin değerleri ve gradient bilgileri temel alınarak bir tanımlayıcı ile temsil edilir. AKAZE özellikleri ölçeğe, rotasyona, sınırlı afinlere göre değişmezdir ve farklı ölçeklerdeki nesnelerde daha fazla ayırt edicilik sağlar.

Bu yerel özellikler arasında seçim yapılırken, özellik sayısı, hesaplama maliyeti ve ölçek, rotasyon ve afin dönüşlere karşı sağlamlık gibi faktörler önemlidir. Tablo 2.6'da görüldüğü gibi, bazı çalışmalarda farklı yöntemlerden elde edilen sonuçları karşılaştırmak [16], [80] ya da farklı yöntemleri bir arada kullanmak [16], [17], [31], [32], [36], [45] için birden fazla yöntem kullanılmıştır

Şekil ve çizgi özellikleri, nesne şekilleri ve sınırları hakkında bilgi sağlar. Harris-Affine [63], Boosted Haar benzeri özellikler [55], HOG [54] ve Hough dönüşümü [53] gibi yöntemler, ürün tanıma çalışmalarında kullanılan şekil ve çizgi özellikleridir.

Harris-Affine [63], afin dönüşlere karşı invaryans olan bir köşe detektörüdür. Köşe noktalarının konumları, ölçekleri ve oryantasyonları ile ilgili bilgiler sağlar.

Boosted Haar benzeri özellikler [55], Haar benzeri özellikleri kullanarak görüntü bölgelerini temsil eder. Her bir Haar özelliği tanımlandıktan sonra elde edilen değer belirli bir eşikten büyükse sınıflandırma sonucu pozitif, aksi takdirde negatif olur. Bu işleme zayıf sınıflandırıcı denir. AdaBoost algoritması kullanılarak elde edilen

en güçlü zayıf sınıflandırıcılar bir araya getirilerek güçlü bir sınıflandırıcı oluşturulur.

HoG [54], bir görüntüdeki gradyan yönlendirmelerinin dağılımını temsil eden bir özellik tanımlayıcısıdır. Görüntüdeki kenar, hat ve desen gibi yapıları temsil eder. Hough dönüşümü [53], görüntülerde basit geometrik şekillerin (çizgiler veya daireler gibi) tespiti için kullanılan bir tekniktir. Kenar tespiti ve şekil analizi için kullanılır.

Şekil ve çizgi özellikleri, diğer özelliklere göre daha az kullanılmış olabilir, ancak bazı çalışmalarda sistem performansını artırmak için renk veya yerel özelliklerle birlikte kullanıldığı görülmektedir [16] [17] [31] [36] [37] [45].

2.3.2.2 Derin Öğrenme Tabanlı Özellik Çıkarma Yöntemleri

Derin öğrenme tabanlı özellikler, geniş ürün görselleri veri kümeleri üzerinde eğitilmiş CNN'ler kullanılarak çıkarılır. Ürün tanıma konusunda daha önce kullanılan derin öğrenme tabanlı özellik çıkarma yöntemleri arasında CaffeNet [65], ResNet [76], AlexNet [64], VGG [74], ResNext [86] ve SNN [62] yer almaktadır. Tablo 2.7, bu derin öğrenme tabanlı özellik çıkarma yöntemlerinin kullanıldığı çalışmaları listelemektedir.

AlexNet [64] ve CaffeNet'in [65] her ikisi de derin bir CNN modelidir. CaffeNet, AlexNet ile benzer bir mimariye sahiptir ve AlexNet'in geliştirilmiş bir halidir [65]. İki model de, evrişimsel katmanlar, tam bağlantılı katmanlar ve son olarak bir softmax çıkış katmanından oluşur. Evrişimsel ve tam bağlantılı katmanlarda rektifiye doğrusal birim (ReLU, Rectified Linear Unit) aktivasyon fonksiyonunu kullanarak modelin daha karmaşık ve lineer olmayan ilişkileri öğrenmesi sağlanır [64]. Nesne tanıma problemlerinde kullanılan bu yöntemler, büyük ölçekli derin öğrenme uygulamaları için bir temel oluşturmuştur ve derin öğrenmenin yükselişine öncülük etmiştir.

ResNet [76], derin öğrenme modellerinin eğitimini kolaylaştırarak daha derin modellerin kullanılabilmesine olanak sağlamıştır. Kullanılan artık bloklar ile ağın bir önceki katmanının çıkışını doğrudan bir sonraki katmanın girişine ekleyen kısa yol bağlantıları içerir. Yaygın olarak kullanılan tam bağlantılı katmanlar yerine küresel ortalama birleştirme kullanarak özellik haritasındaki her bir özellik için ortalama alarak boyutsuz bir vektör elde etme avantajına sahiptir. Batch

normalizasyonu kullanarak ağın eğitimini hızlandırır ve aşırı uydurmayı azaltır. Böylece daha istikrarlı bir eğitim süreci sağlar. ResNet'in farklı derinliklerdeki modellerinin kullanıldığı ResNet-18, ResNet-50, ResNet-101 ve ResNet-152 gibi çeşitleri mevcuttur. ResNet mimarisinin geliştirilmiş bir uzantısı olan ResNext [86], gruplandırılmış evrişim kavramını içerir. Gruplanmış evrişimlerdeki grup sayısını temsil eden kardinalite fikri ile paralel olarak çalışan birden çok küçük ağı, tek bir büyük ağı tercih edilebileceğini savunur [82].

Tablo 2.3 Derin öğrenme tabanlı özellik çıkarma yöntemleri

Derin Öğrenme Tabanlı Özellik Çıkarma Yöntemleri	Yöntemlerin Kullanıldığı Kaynaklar
CaffeNet [65]	[46]
ResNet [76]	[81] [48]
AlexNet [64]	[15]
VGG-16 [74]	[47]
ResNext [86]	[40]
SNN [62]	[42]

VGG [74], basitliği ve etkinliği ile bilinen bir CNN mimarisidir. Temel yapısı, ardışık olarak birbirine eklenen evrişimli katmanlar, tam bağlantılı ve maksimum havuzlama katmanlarından oluşur. 16 veya 19 katmanlı olabilen iki farklı sürümü vardır.

SNN [62], benzerlik ölçümü için tasarlanmış bir çift ağı yapısını temsil eder. İki benzer görüntüyü girdi olarak alır ve bu görüntüler arasındaki benzerliği ölçen bir çıkış üretir. Giriş örnekleri üzerindeki öğrenilmiş temsiller arasındaki benzerlik, genellikle öklidyen uzaklık veya benzer bir ölçümle belirlenir. Kullanılan kontrastif kayıp fonksiyonu, aynı sınıfa ait örnekler arasındaki benzerliği artırmayı ve farklı sınıflara ait örnekler arasındaki benzerliği azaltmayı sağlar. Siyam ağlarının benzerlik ölçümü gerektiren uygulamalar için kullanımı uygundur.

Bu modeller, derin öğrenme tabanlı özellik çıkarma konusunda çeşitli başarılar elde etmiş ve geniş bir uygulama yelpazesi için kullanılmıştır. Derin öğrenme tabanlı özellik çıkarma yöntemlerinin market raflarındaki görüntülerin tanınması için kullanılmasının çeşitli avantajları ve dezavantajları olabilir. Birbirine çok benzer ürün örneklerinin tanınması için derin yöntemler daha ayırt edici özellikleri yakalayabilir. Ayrıca, derin öğrenme modelleri, bilgisayarla görme yöntemleri kullanılarak özellik çıkarılırken karşılaşılan çeşitli görsel değişikliklere (ışık koşulları, perspektif değişiklikleri, vb.) karşı dayanıklılık gösterebilir. Öte yandan, derin öğrenme modelleri büyük ve temsil edici veri setlerine ihtiyaç duyar. Eğer yeterli ve çeşitli veri sağlanamazsa, modelin özellik çıkarma yeteneği düşebilir. Market raflarında ürün tanıma probleminde ürün çeşidinin fazla olması ve sürekli eklenen yeni ürünler ya da sıklıkla ambalajı değişen ürünlerin olması nedeniyle veri elde etme ve etiketleme süreci uzun ve maliyetlidir. Bu nedenle özellik seçiminde hangi modellerin kullanılabileceği ilgili görevlere ve veri setlerine göre değişebilir.

2.3.3 Özelliklerin Sınıflandırılması için Kullanılan Yöntemler

Ürünün sınıfını veya konumunu tespit etmek için elde edilen özelliklerin ürünün şablon görselleriyle eşleştirilerek sınıflandırılması gerekmektedir. Tablo 2.8'de listelendiği üzere, bunun için farklı geleneksel ve derin sınıflandırma yöntemleri kullanılmıştır. Her biri farklı bir sınıflandırma yaklaşımını temsil eden algoritmalar: hiyerarşik k-ortalamlar [87], SVM, VGG-16 [74], VGG-f [71], Hungarian algoritması [79], Naive Bayes sınıflandırıcısı [88], rastgele ormanlar [89] ve ELM'dir [57].

Hiyerarşik k-ortalamlar [87], k-ortalamlar algoritmasının geliştirilmiş halidir. Kümeleme işlemini ağaca benzeyen hiyerarşik bir yapı içinde gerçekleştirir. Özellikleri sınıflandırmak için görüntülerdeki benzer özellikleri bir araya toplar. Bunun için öncelikle, özellikleri küçük alt gruplara ayırır, sonra her alt grup içinde k-ortalamlar uygulayarak alt kümeleri belirler.

SVM, veri noktalarını farklı sınıflara en iyi şekilde ayıran hiper-düzlemi bularak çalışan, denetimli bir makine öğrenimi algoritmasıdır. Bu hiper-düzlem, sınıflar arasındaki marjı maksimize etmeye çalışır. Özellikle düşük boyutlu veri setlerinde etkilidir ve genelleme yeteneği iyidir.

Tablo 2.1 Özelliklerin sınıflandırılması için kullanılan yöntemler

Özelliklerin Sınıflandırılması için Kullanılan Yöntemler	Yöntemlerin Kullanıldığı Kaynaklar
Hiyerarşik K-Ortalamalar [87]	[45]
SVM	[17] [34] [37] [50]
VGG-16 [74]	[80]
VGG-f [71]	[33]
Hungarian Algoritması [79]	[49]
Naive Bayes Sınıflandırıcısı [88]	[32]
Rastgele Ormanlar [89]	[6]
ELM [57]	[39]

VGG-16 [74] ve VGG-f [71] derin evrişimli sinir ağı mimarileridir. VGG-16, 16 katmandan oluşurken, VGG-f onun daha hafif bir çeşididir.

Sıralama veya eşleme tablolarında kullanılan Hungarian algoritması [79], iki görüntü arasındaki özellikleri eşleştirir. Bunun için, iki farklı sıralama veya eşleme arasındaki optimal eşleme problemlerini çözer. Çift yönlü ağırlıklı bir grafik oluşturur ve minimum ağırlıklı eşlemeyi bulunmaya çalışır.

Naive Bayes sınıflandırıcısı [88], Bayes teoremine dayalı olarak özellikler arasında bağımsızlığı varsayan olasılıksal bir sınıflandırma algoritmasıdır. Hızlı ve basit bir sınıflandırma algoritmasıdır. Fakat özellikler arasındaki bağımsızlık varsayımı gerçek dünyadaki durumları tam olarak yansıtmayabilir.

Rastgele ormanlar [89], eğitim sırasında bir dizi karar ağacı oluşturarak eğiten ve sınıfları tahmin olarak çıkaran bir topluluk öğrenme yöntemidir. Aşırı öğrenmeye karşı dayanıklı bir algoritmadır.

ELM [57], bir giriş katmanı ve bir çıkış katmanından oluşur. Ağırlıklar rastgele atanır ve ardından çıkış katmanındaki ağırlıklar analitik olarak hesaplanır. ELM’de

eğitim süreci hızlıdır. Ayrıca, doğrusal ve doğrusal olmayan sınıflandırma görevlerinde kullanılabilir.

Bu algoritmalar, belirli kullanım senaryolarına ve veri setlerine bağlı olarak avantajlı olabilir. Her bir algoritmanın güçlü ve zayıf yönlerini anlamak, uygun algoritmayı seçmeye yardımcı olur.

2.3.4 Sonuçların İyileştirilmesi için Kullanılan Yöntemler

İyileştirme aşaması, raf görüntüsünde bulunan her ögeye karşılık gelen birden fazla üst üste bindirilmiş sınırlayıcı kutuyu tek bir sınırlayıcı kutuda birleştirerek nihai sonuca ulaşmayı amaçlamaktadır. Tablo 2.9'da listelendiği gibi, bunun için maksimum olmayan bastırma (NMS, Non-Maximum Suppression) [90], kümeleme algoritmaları, grafik tabanlı yaklaşımlar, saklı Markov Modelleri (HMM, Hidden Markov Models) ve K-en yakın komşu (K-NN, K-Nearest Neighbour) gibi algoritmalar kullanılmıştır.

NMS [90], ürün tanıma probleminin iyileştirme aşamasında çokça kullanılan bir yöntemdir. Aynı nesne için birden fazla sınırlayıcı kutu önerisi sorununu giderir. NMS sırasında, sınırlayıcı kutular güven puanlarına göre sıralanır ve daha düşük puanlara sahip çakışan kutular bastırılır ve yalnızca en yüksek güvene sahip kutu tutulur. Böylece, nesne tespiti modellerinin çıkışında ortaya çıkan çoklu çerçeveler temizlenerek sistem doğruluğu artırılır.

Farklı özelliklerdeki kümeleme algoritmaları da aynı ürün için çoklu bölge önerilerini azaltmak için kullanılmaktadır. Benzer özelliklere sahip sınırlayıcı kutuları aynı gruplar halinde düzenleyerek ürün konumu doğruluğunu artırır [15].

Ürün tanıma bağlamında grafik tabanlı yaklaşımlar genellikle raflarda yer alan ürünler arasındaki ilişkilerin bir grafik olarak temsil edilmesini ve tanıma sonuçlarını iyileştirmek için grafik tabanlı algoritmalarla yararlanmayı içerir. Aynı ürün için birden fazla sınırlayıcı kutu önerisini bu grafik tabanlı modelden elde edilen bilgilere göre indirger.

HMM'ler, zamanla değişen bir sürecin ardındaki gizli durumları modellemek için kullanılır. Gözlemlenen veriler, gizli durumların olasılıklarını etkiler. Ürünlerin belirli bir düzen içinde dizildiği durumları modelleyerek, Aynı ürün için birden fazla sınırlayıcı kutu önerisini bu modelden yararlanarak indirger [16].

K-NN sınıflandırıcısı, bir veri noktasını sınıflandırmak veya tahmin etmek için en yakın komşularını kullanır. Ürün tanımadaki iyileştirme aşamasında, sınırlayıcı kutular içinden en benzer k örneği seçer ve bunları tek bir sınırlayıcı kutuya indirger [47].

Tablo 2.1 Sonuçların iyileştirilmesi için kullanılan yöntemler

Sonuçların İyileştirilmesi için Kullanılan Yöntemler	Yöntemlerin Kullanıldığı Kaynaklar
NMS [90]	[15] [33] [81]
Kümeleme Algoritmaları	[15]
Grafik Tabanlı Yaklaşımlar	[16]
HMM	[34]
K-NN	[47]

Bu yöntemler, özellikle aynı nesne için birden fazla sınırlayıcı kutu teklifinin oluşturulduğu senaryolarda, ürün tanıma ve yerelleştirme sistemlerinin çıktısını iyileştirmek için yaygın olarak kullanılır. İyileştirme teknikleri, tespit edilen nesnelerin hassasiyetini ve güvenilirliğini artırmak ve nihai sonuçlarda yalnızca en güvenilir ve doğru tahminlerin korunmasını sağlamak için çok önemlidir. Belirli bir iyileştirme yönteminin seçimi, hesaplama verimliliği, ölçeklenebilirlik ve tespit görevinin özellikleri gibi faktörlere bağlıdır.

3

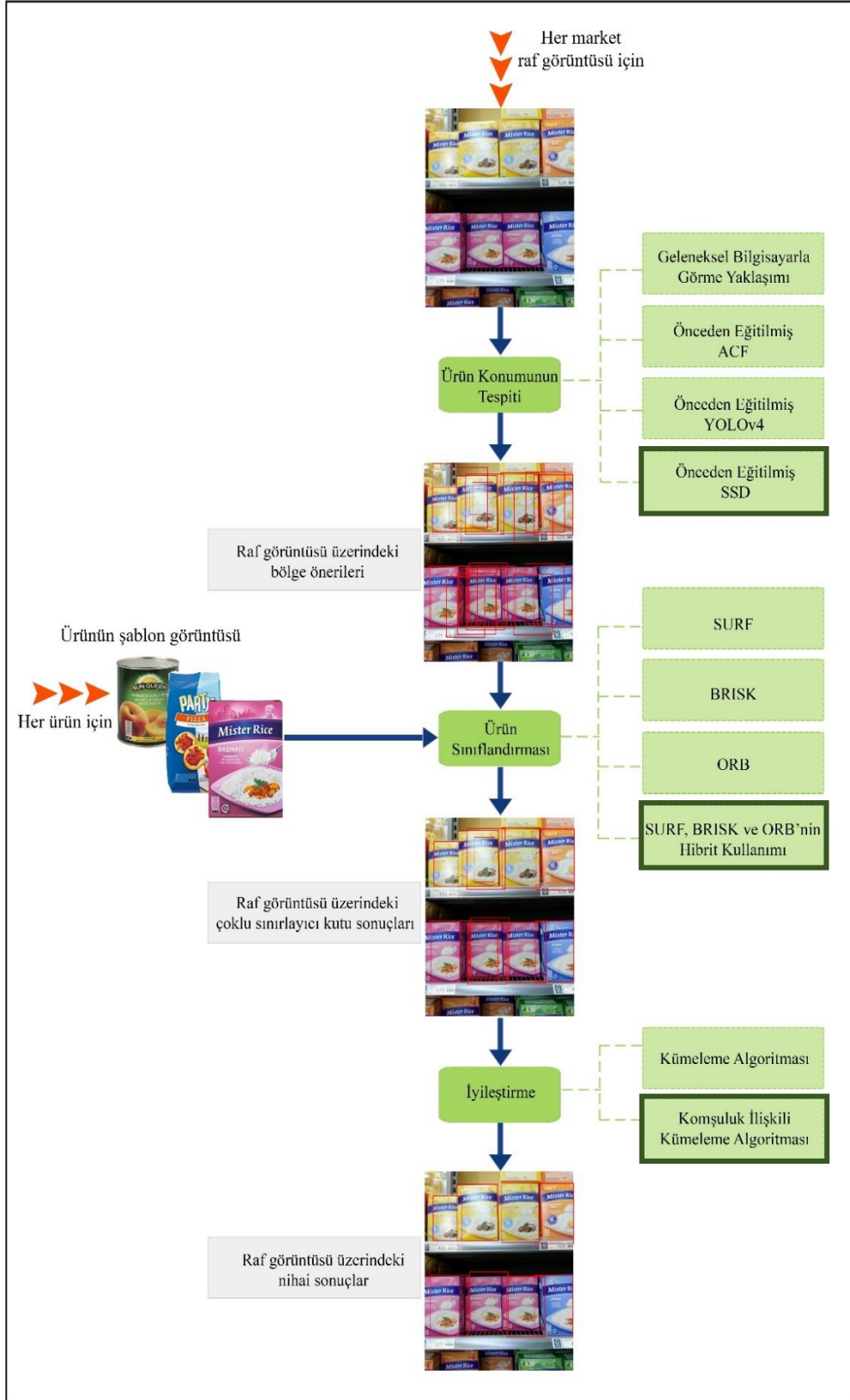
ÇOK AŞAMALI UÇTAN-UCA ÜRÜN TANIMA SİSTEMİ

3.1 Önerilen Yöntem

Market raf görüntülerinde ürün tanıma problemi, çok sınıflı sınıflandırma, veri sınırlılığı, sınıf içi çeşitliliğin fazla olması ve sistemin yeni ürünlere, farklı raf tasarımlarına esneklik gösterebilmesi gibi zorluklar nedeniyle klasik bir nesne tanıma probleminden daha karmaşıktır [72]. Bu zorlukların üstesinden gelmek amacıyla, bu tez çalışmasında problemin farklı aşamaları için karşılaştırılabilir sonuçlara sahip farklı çözümler önerilmektedir.

Şekil 3.1'de bu tez çalışmasında önerilen çok aşamalı uçtan-uca ürün tanıma yaklaşımı gösterilmiştir. Ürün tanıma süreci üç aşamadan oluşmaktadır: (1) ürün tespiti, (2) ürün sınıflandırması ve (3) iyileştirme aşamaları. Ürün tanıma sürecinin her aşaması için birden fazla yöntem önerilmiş ve hangi yöntemlerin daha başarılı sonuçlar elde edeceği deneysel olarak araştırılmıştır. Bu tez çalışmasında önerilen çok aşamalı yaklaşım ile ürün tanımada karşılaşılan ve aşağıda listelenen zorlukların çözümüne önemli katkılar sağlanacağı iddia edilmektedir:

- Son teknoloji nesne algılama yöntemlerinin (SSD [9], Faster R-CNN [28], YOLO'nun farklı versiyonları [29], [59], [8], [75], [91]) performansları oldukça gelişmiş olmasına rağmen, sınıf sayısı arttıkça nesne tanıma performansı düşmektedir [72]. Bu nedenle, son teknoloji nesne tespit yöntemleriyle market ürünlerinin doğrudan tanınması, çok sınıflı veri kümeleri için yeterli değildir.
- Market raflarında ürün tanıma problemi veri toplama, sınıf sayısının fazla olması, sürekli yenilenen ambalajlar ve yeni eklenen ürünlerin varlığı nedeniyle zorlu bir problemidir. Veri setlerindeki bu sınırlamaların üstesinden gelebilmek için sınıflara ait görsel sayısının yetersiz olduğu durumlarda, tek sınıf üzerinden nesne tespit algoritması ile ürünlerin konumları belirlenebilir ve daha sonra ürün tanıma ile ürünler sınıflandırılabilir [92], [72].



Şekil 3.1 Çok aşamalı uçtan-uca ürün tanıma süreci
(Her aşama için en iyi sonucu yöntem koyu renk ile vurgulanmıştır.)

- Sürekli yenilenen ambalajların ve yeni eklenen ürünlerin varlığı, doğrudan ürün tanıma uygulandığında sistemin yeniden eğitilmesini gerektirmektedir. Bu tez çalışmasında önerilen üründen bağımsız ürün tespit süreci, marketlerde eklenen veya çıkarılan ürünlerden etkilenmemekte ve daha esnek tespit imkânı sağlamaktadır.

Bunlara ek olarak, bu tez çalışmasında önerilen hibrit yöntem ile ürün sınıflandırma aşamasında aşağıda belirtilen zorlukların üstesinden gelineceği iddia edilmektedir:

- Ürünler arasındaki benzerlik, ürün boyutlarının farklı olması, renk ve şekil çeşitliliği tüm ürünlerin tek bir yöntemle yeterince tanınamamasına yol açmaktadır. Bu nedenle, daha geniş bir ürün sınıfı yelpazesini başarılı bir şekilde tanımak için birden fazla özellik türüne ortaklaşa güvenmek gerekir. Bu amaçla, bu çalışmada SURF [10], BRISK [11] ve ORB [14] özelliklerinin hibrit kullanımı önerilmiştir.

Önerilen çok aşamalı uçtan-uca ürün tanıma yöntemi Şekil 3.1'de gösterilmektedir. Gösterildiği üzere, sistem girdi olarak bir raf görüntüsünü alır. Raf görüntüsünde yer alan ürünlerin potansiyel bölgelerini bulur. Ardından ürünlerin şablon görüntüsü ile bulunan potansiyel bölgeler arasındaki özellikleri eşleştirir ve son olarak aynı ürün için bulunan birden fazla tespit ve sınıflandırmayı tek bir sonuca indirger.

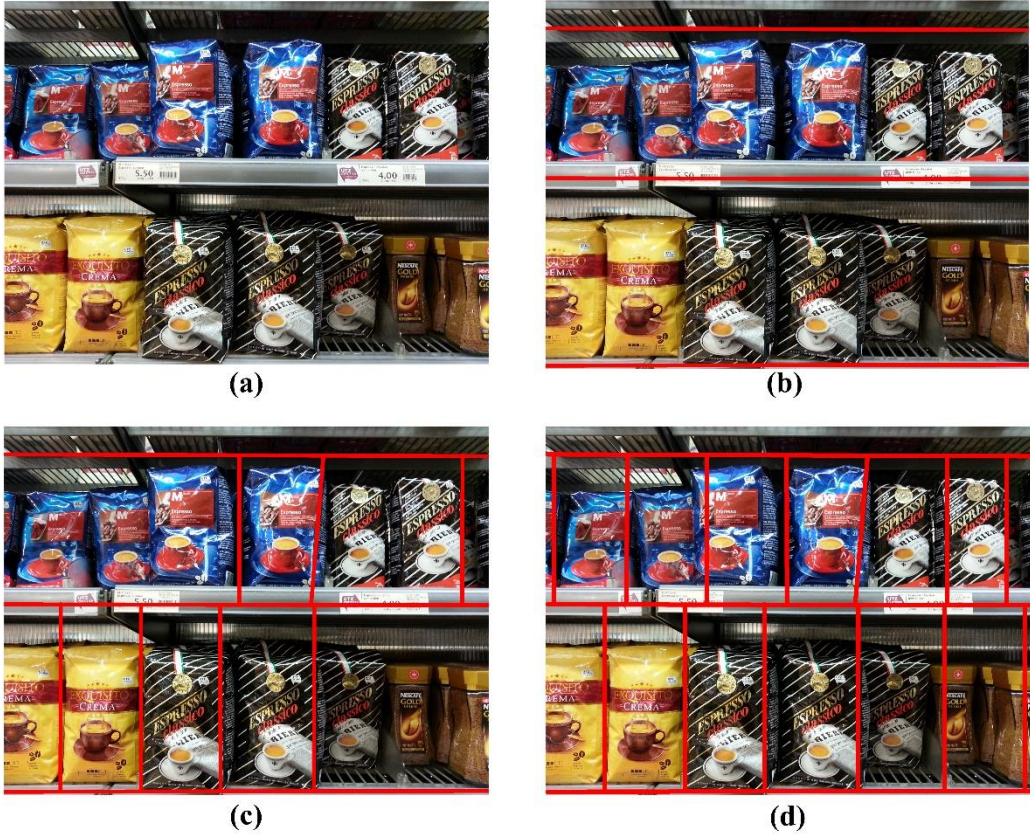
3.1.1 Ürün Tespiti Aşaması

Market raflarında ürün tanıma problemi için önerilen üç aşamalı yöntemin ilk aşaması olan nesne tespitinde raflardaki her bir mevcut ürünün potansiyel bölgeleri belirlenir. Bu tez çalışmasında ürün tespit aşaması için dört farklı yöntem kullanılır ve bu yöntemlerin performansları karşılaştırılır. Bu yöntemlerden ilki geleneksel bilgisayarla görme yaklaşımı, ikincisi öğrenmeye dayalı bir ACF [7] detektörü ve son olarak DNN tabanlı YOLOv4 [8] ve SSD [9]. Farklı yöntemleri kullanmanın ve karşılaştırmanın amacı, bu üç farklı temeldeki yaklaşımın ürün tanıma problemi performansındaki avantajları ve dezavantajlarını göstermektir.

3.1.1.1 Geleneksel Bilgisayarla Görme Yaklaşımı

Bilgisayarla görme yaklaşımı için market raf görüntülerinin genel özelliklerinden faydalanılır. Market görüntüleri farklı sayıda raflardan ve farklı sayıda ürünlerden oluşur. Konumları tespit etmek istenen ürünler bu raflar arasında yer alır. Bu

nedenle, görüntü üzerinde ürünlerin aranacağı alanı sınırlamak için önce raf çizgileri sonrasında tespit edilen raf çizgilerinin arasında ürünler aranır.



Şekil 3.1 Aşama-1'de önerilen geleneksel bilgisayarla görme yaklaşımının adımları: (a) orijinal raf görüntüsü; (b) ilk adımdan sonra tespit edilen raf çizgilerini içeren raf görüntüsü; (c) ikinci adımdan sonra tespit edilen ürün bölgelerini içeren raf görüntüsü; (d) üçüncü adımdan sonra tamamlanmış ürün bölgelerini gösteren raf görüntüsü

Önerilen geleneksel bilgisayarla görme yaklaşımı Şekil 3.2'de detaylandırıldığı üzere üç adımdan oluşmaktadır. İlk olarak, rafların tespiti için Hough Dönüşümü [53] ile raf görüntüleri üzerindeki yatay çizgiler belirlenir; ardından, ürünlerin her bir raftaki başlangıç ve bitiş noktalarını bulmak için yine Hough Dönüşümü [53] ile dikey çizgiler vurgulanır. Bu sürecin sonucunda, bazı ürünlerin başlangıç ve bitiş noktaları kaçırılıp belirlenemeyebilir. Bu durum çalışmanın daha ilk aşamasından birçok ürün konumunun tespit edilememesine ve sistem performansının düşmesine neden olur. Bu sorunun çözümü için bir iyileştirme adımı sunulmuştur. Eksik ürün konumları, önceki adımlardan elde edilen alan bilgisi kullanılarak tamamlandı. Raflarda birbiri ile aynı ya da benzer boyutlardaki ürünlerin yan yana olma

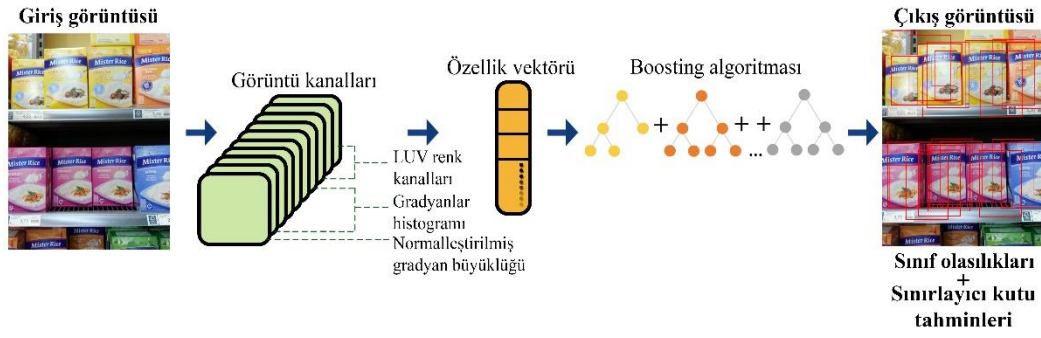
olasılığının yüksek olduğu bilgisinden yararlanılarak tespit edilen alanlar üzerinden ek bölge önerileri oluşturulmuştur.

3.1.1.2 ACF Detektörü

Market raflarında ürün tanıma probleminin nesne tespiti aşaması için önerilen ikinci yöntem olarak öğrenme tabanlı ACF [7] detektörü kullanılmıştır. Şekil 3.3'te akış diyagramı verilen detektör, görüntüden elde edilen özelliklerin sınıflandırılması ile çalışmaktadır. Giriş görüntüsünü bir özellik vektörü olarak temsil etmek için normalleştirilmiş gradyan büyüklüğü, yönlendirilmiş gradyanların histogramı (altı kanal) ve LUV renk kanalları (toplam: 10 kanal) hesaplanır ve ardından görüntünün sınıflandırılması için bir Boosting algoritması [93] kullanılır.

Eğitim sırasında, her biri bir ürünü çevreleyen, ürünün çeşitli boyutlarını ve konumlarını temsil eden sınırlayıcı kutulardan oluşan raf görüntüleri kullanılır. Her bir sınırlayıcı kutu alanı üzerinden elde edilen özellik vektörleri Boosting algoritması kullanılarak bir dizi zayıf sınıflandırıcıyı yinelemeli olarak eğitir. Her zayıf sınıflandırıcı, her yinelemede yanlış sınıflandırılmış örneklerle daha fazla ağırlık atanarak eğitim örneklerini sınıflandırmak üzere eğitilir. Her zayıf sınıflandırıcının performansı değerlendirilir ve eğitim verileri üzerinde en iyi performansı gösteren sınıflandırıcıya, son birleştirilmiş sınıflandırıcıda daha fazla ağırlık verilir. Bu sayede, daha iyi performans gösteren sınıflandırıcılar daha fazla etkiye sahip olur. Bu süreç, sonraki zayıf sınıflandırıcıların önceki zayıf sınıflandırıcılar tarafından yanlış sınıflandırılan örneklerle daha fazla odaklanmasıyla yinelemeli olarak devam eder. Eğitilmiş zayıf sınıflandırıcıların ağırlıklı oylaması yoluyla elde edilen ve tüm zayıf sınıflandırıcıların ortak kararını temsil eden son birleştirilmiş sınıflandırıcı, yeni raf görüntülerinde çok ölçekli bir kayan pencere yaklaşımı ile oluşan alanlar üzerinde ürün konumlarının tespiti ve sınıflandırılması için kullanılır.

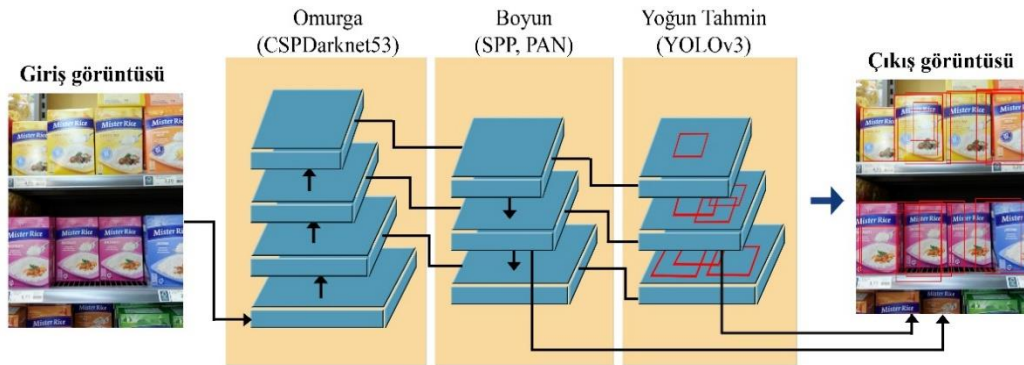
ACF ve ACF'nin farklı uygulamaları, insanların [94], çekirgelerin [95], uçakların [96] ve ultrason görüntüsünde fetal uzvun kesit alanlarının [97] tespiti gibi bazı nesne algılama problemlerinde iyi performans elde etmiştir. ACF daha önce ürün tanıma için kullanılmamış olsa da benzer özelliklere sahip nesnelerin tanınmasında iyi bir performansa sahip olması nedeniyle, birbirine benzer ürünler için bölge önerileri oluşturulurken de yeterli bir performans gösterebilmesi beklenmektedir.



Şekil 3.2 Aşama-1'de önerilen ACF detektörünün adımları

3.1.1.3 YOLOv4

Nesne tespiti aşaması için önerilen DNN tabanlı YOLOv4'un [8] yapısı Şekil 3.4'te gösterilmektedir. Burada raf görüntüleri giriş görüntüsü olarak sisteme verilir. Modelin omurga (backbone) kısmı, özellik çıkarımının yapıldığı katmandır. Özellik çıkarım modeli olarak CSPDarknet53 kullanılır. Omurga ile yoğun tahmin (bu kısım için baş-head terimleri de kullanılır) kısmı arasında yer alan kısımda boyun (neck) olarak adlandırılır. Modelin boyun kısmında kısmi nesnelerin tahmini yapılırken daha fazla bilgi elde edilmesi için kullanılır. Burada FPN kullanılarak aşağıdan yukarı ya da yukarıdan aşağı akıştan gelen bilgilerle komşu özellik haritalarından detaylı bilgi çıkarımları yapılır. YOLOv3'te [59] kullanılan FPN yerine YOLOv4'te [8] yol toplama ağının (PAN, Path Aggregation Network) değiştirilmiş bir sürümü ve modifiye edilmiş uzamsal piramit havuzu (SPP, Spatial Pyramid Pooling) kullanılır. Son olarak yoğun tahmin kısmında da YOLOv3'te [59] olduğu gibi sınırlayıcı kutular tespit edilir ve her kutunun sınıfı tahmin edilir.



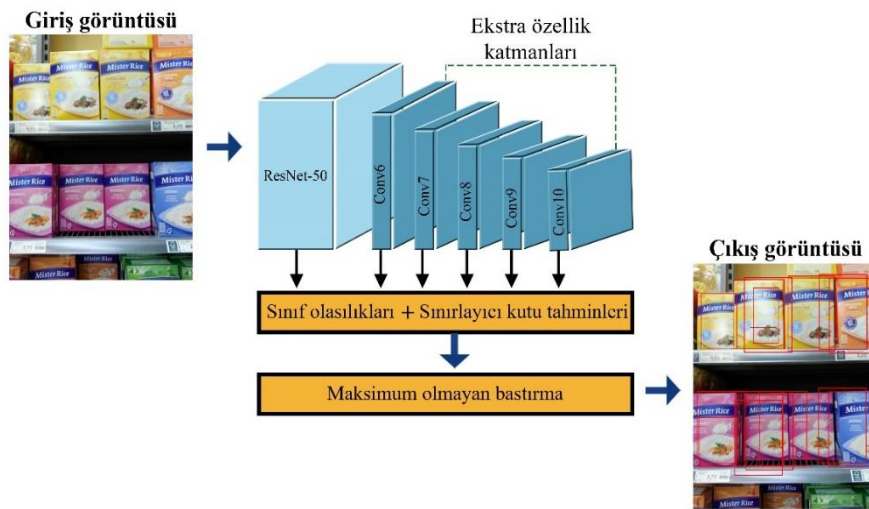
Şekil 3.3 Aşama-1'de önerilen YOLOv4 detektörünün adımları

Önerilen yeni özellik çıkarıcı modeli CSPDarknet53 ve hızlı algılama doğruluğu sağladığından dolayı Aşama-1'in DNN tabanlı modeli olarak tercih edilmiştir.

YOLOv4 [8] kullanılan modeli eğitmek için önceden eğitilmiş ImageNet ağırlıkları kullanılmıştır.

3.1.1.4 SSD

Nesne tespiti aşaması için önerilen son yöntem olarak, DNN tabanlı SSD kullanılmıştır. Regresyon ve sınıflandırma için eşzamanlı sınırlayıcı kutu ile temsil edilen tek aşamalı bir detektör olan SSD'nin [9] akış diyagramı Şekil 3.5'te verilmiştir. Tek aşamalı modeller [9] ve [59], nesneleri bölge öneri algoritmalarıyla tespit eden ve ardından her bölge için bağımsız bir sınıflandırma uygulayan iki aşamalı modellerden [28], [98] daha hızlıdır. Diğer tek aşamalı model olan YOLO'nun görüntülerdeki daha küçük nesneleri tanımlama konusunda bir sınırlaması vardır ve alışılmamış en boy oranlarına veya konfigürasyonlara sahip nesneleri genelleştirmesi zordur [99]. Ürün tanıma problemi, farklı ölçek ve ürün boyutlarındaki ürünlerin tanınmasını gerektirmektedir. SSD [9], farklı derinlikteki özellikleri (çok boyutlu özellik haritaları) kullanarak küçük ve büyük nesneleri aynı anda algılayabilir. Ek olarak SSD [9], düşük çözünürlüklü giriş görüntülerinde bile yüksek doğruluk elde eder. Sonuç olarak, belirtilen bu avantajlardan dolayı Aşama-1'in DNN tabanlı bir diğer modeli için de SSD [9] seçilmiştir. ImageNet veri kümesinde ile önceden eğitilmiş ağırlıklar, ResNet-50 [76] temel ağıyla SSD için başlangıç ağırlık modeli olarak kullanılmıştır. Şekil 3.5'te gösterildiği gibi, Aşama-1'in potansiyel ürün bölgeleri, ResNet-50 ağından ve ekstra özellik katmanlarından gelen tahminleri filtrelemek için maksimum olmayan bastırma kullanılarak elde edilir.



Şekil 3.4 Aşama-1'de önerilen SSD adımları

3.1.2 Ürün Sınıflandırma Aşaması

Market raflarında ürün tanıma problemi için önerilen üç aşamalı yöntemin ikinci adımı olan ürün sınıflandırma aşamasında, nesne tespiti aşamasından elde edilen bölge önerilerinin hangi ürün sınıfına ait olduğu tahmin edilir. Raf görsellerinden elde edilen bölge önerileri ile ürünlerin şablon görselleri arasında özellik eşleştirmesi yapmadan önce bölge önerisi sayısının azaltılması için ön-eleme işlemi gerçekleştirilir. Ön-eleme işleminin sözde kodu (pseudocode) Şekil 3.6'da gösterilmektedir. Ön-eleme süreci, gereksiz bölge önerilerini belirtilen iki şekilde ortadan kaldırır: (1) en boy oranı ve (2) renk histogramı. Bu bölgeleri, önerilen yöntemin sonraki aşamalarına dâhil etmez.

Öncelikle, ürün şablonunun ve bölge önerilerinin en boy oranı eleme işlemi için hesaplanır. Burada ürün şablonunun en boy oranının 1.5 katı üst eşik, 0.5 katı ise alt eşik olarak kabul edilir. Bölge teklifinin üst eşikten yüksek veya alt eşikten düşük olması durumunda bu bölge teklifi elenir. Daha sonra, en boy oranı kısıtlamasıyla elenmeyen şablon görüntülerinin normalleştirilmiş HSV renk histogramı çıkarılır. Yaygın olarak kullanılan HSV renk uzayında renk özü (hue), kırmızı, sarı, yeşil ve mavi gibi algılanan renkler arasında ayırım yaparak renk türünü 0 ile 360 arasında değişen bir değerle ifade eder. Doygunluk (saturation) renklerin solgunluğunu veya canlılığını belirtir. Değerler genellikle 0 ile 100 arasında ölçülür. 0 değeri, renksiz veya gri tonu temsil ederken, 100 değeri renklerin en doymuş hallerini ifade eder. Parlaklık (value) değeri ise, 0 (siyah) ile 100 (beyaz) arasındaki renk aralıklarının açıklığını yani ne kadar parlak veya koyu olduğunu temsil eder [84]. Ürün şablonu görselleri ile bölge önerilerinin histogram kesişimi önceden belirlenmiş bir eşik değerinden düşükse bu bölgeler elenir. Böylece, geri kalan bölge önerileri sonraki adımlar için yeni bölge öneri kümesini oluşturur. Ön elemelemenin avantajı, özellik çıkarma adımından önce önerilen bölge sayısını azaltarak sonraki adımlar için yapılacak işlem sayısını azaltmaktır.

Ön-eleme işleminden sonra, her bir ürünü sınıflandırmak için ürün şablonlarının ve geri kalan bölge önerilerinin özellikleri çıkarılır. Literatürde yaygın olarak kullanılan özellik çıkarma algoritmaları incelendiğinde maksimum kararlı ekstrem bölgeleri özellikleri (MSERF, Maximally Stable External Region Features) [100], AKAZE [85], SURF [10], BRISK [11] ve ORB [14] özelliklerinin öne çıktığı görülmektedir. Bu özellik çıkarma yöntemlerinin market raf görüntülerinde ürün

tanıma problemi için uygunluna bakılmıştır. Tablo 3.1’de görüldüğü üzere MSERF [100] ve AKAZE [85] ile GP-20 [15] veri setindeki ürünlere ait şablon görüntülerden elde edilen özellik sayısı diğer özellik çıkarma yöntemlerinden oldukça azdır. Şablon görüntülerden az özellik çıkarılması, ileriki aşamalarda market raf görüntülerinden elde edilen aday pencereler üzerinde daha da az özellik çıkarılması anlamına gelmektedir. Çünkü şablon görüntüler ideal koşullarda çekilen görüntülerden oluşurken, market raf görüntüleri ortam koşullarından (parlama, yansıma, farklı açılardan çekilme, vb.) etkilenen görüntülerdir.

```

function: on_eleme_islemi (T,R,Thr_Th,Thr_Tl,Thr_ch)

  Input: T = şablon görüntü kümesi
           R = bölge önerileri görüntü kümesi
           Thr_Th = şablon görselinin en yüksek en-boy oranı
           Thr_Tl = şablon görselinin en düşük en-boy oranı
           Thr_ch = renk histogramlarının kesişimi için eşik değeri

  Output: R_new = ön eleme süreci sonrasındaki yeni bölge önerileri

  R_new ← ∅

  for each Ti
    [MT, NT] ← Ti 'nin boyutu
    Thr_Th ← max(MT, NT) / min (MT, NT) * 1.5
    Thr_Tl ← max(MT, NT) / min (MT, NT) * 0.5
    HTHSV ← Ti 'nin renk histogramının hesaplanması
    for each Rj
      [MR, NR] ← Rj 'nin boyutu
      if max(MR, NR) / min (MR, NR) < Thr_Th || max(MR, NR) /
      min (MR, NR) > Thr_Tl then
        HRHSV ← Rj 'nin renk histogramının hesaplanması
        if (HRHSV ∩ HTHSV) > Thr_ch
          R_new ← R_new ∪ Rj
        end if
      end if
    end for
  end for

  return R_new

```

Şekil 3.1 Aşama-2'de önerilen ön-eleme işleminin sözde kodu

Ayrıca, özellik çıkarma yöntemlerinin karşılaştırmalı analizinin yapıldığı çalışma [101]'e göre de, SURF [10], BRISK [11] ve ORB [14] verimli hesaplama performansı ile en yüksek sayıda ayırt edici özelliği çıkarmaktadır. Ek olarak, SURF ve BRISK ölçek açısından en değişmez özellik detektörleridir. BRISK ve ORB de diğerlerine göre daha fazla dönme açısından değişmezdir [101]. Bu nedenle çalışmaya, şablon görüntülerde en çok özellik elde eden SURF, BRISK ve ORB özellikleri ile devam edilmiştir. SURF, BRISK ve ORB'nin özellik çıkarma işlemleri Şekil 3.7'de de gösterildiği üzere üç ana adımdan oluşur: (1) anahtar nokta tespiti, (2) özellik vektörü tanımlaması ve (3) farklı görüntüler arasındaki vektörlerin eşleştirilmesi.

3.1.2.1 SURF

Yaygın olarak kullanılan ölçekle değişmeyen özellik çıkarma yöntemi SURF'te [10], eşleştirilecek iki görüntünün anahtar noktaları yatay ve dikey olarak Gauss yumuşatma filtreleri ile yuvarlak şekilli yapılar olarak tespit edilir; daha sonra her bir anahtar nokta için Hessian matrisine entegre edilir. Sonrasında her kilit noktanın komşuluğu 4×4 'lük bir dizi alt kare bölgeye bölünür. Giriş görüntüsü ve onun integral görüntüsü için iki boyutlu Haar dalgacığının hesaplanmasıyla tüm alt bölgelerin 64 boyutlu ($4 \times 4 \times 4$) özellik vektörü elde edilir. Eşleştirme adımında iki görüntünün özellik vektörleri arasındaki uzaklık Öklid mesafesi kullanılarak hesaplanır. İki özellik vektörü, aralarındaki uzaklık eşik değerinden az olduğunda eşleştirilir.

3.1.2.2 BRISK

Uyarlanabilir ve genel hızlandırılmış segment testini (AGAST, Adaptive and Generic Accelerated Segment Test) [102] kullanarak köşeleri tespit eden ve bunları FAST'ten [12] gelen köşe puanıyla filtreleyen BRISK'de [11] ölçeğe değişmezlik sağlayan ölçek uzay piramidindeki maksimumlar aranır. Ölçek uzay piramidindeki maksimum, ölçeğe değişmezlik sağlar. BRISK tanımlayıcısı [11], anahtar noktaların yoğunluk değerlerinin karşılaştırılmasına dayanmaktadır. Birinci noktanın değeri ikinci noktadan büyük olduğunda çıkış değeri 1, aksi durumda 0 olur. Böylece 512 bit uzunluğunda bir tanımlayıcı elde edilir. Eşleştirme adımında Hamming mesafesi kullanılır, bu da yürütme süresinin Öklid mesafesine kıyasla

daha kısa olmasını sağlar. Son olarak, iki özellik vektörü arasındaki mesafe sabit bir eşikten az ise tanımlayıcılar eşleştirilir.



Şekil 3.2 Aşama-2’de önerilen ürün sınıflandırma aşamaları

Tablo 3.1 Farklı özellik çıkarma yöntemleri ile GP-20 veri setindeki ürün şablon görüntülerinden elde edilen özellik sayıları.

GP-20 Veri Seti Ürün Şablonları	MSERF	AKAZE	SURF	BRISK	ORB
Ürün Şablonu 1	35	78	306	353	412
Ürün Şablonu 2	51	145	382	465	484
Ürün Şablonu 3	56	168	413	386	655
Ürün Şablonu 4	42	122	522	624	603
Ürün Şablonu 5	87	129	513	782	706
Ürün Şablonu 6	54	175	391	453	621
Ürün Şablonu 7	92	142	583	768	816
Ürün Şablonu 8	36	72	163	145	223
Ürün Şablonu 9	55	167	403	462	501
Ürün Şablonu 10	31	62	146	213	281
Ürün Şablonu 11	23	86	119	189	256
Ürün Şablonu 12	86	121	595	662	853
Ürün Şablonu 13	72	206	290	389	557
Ürün Şablonu 14	56	121	566	724	932
Ürün Şablonu 15	62	211	465	716	857
Ürün Şablonu 16	46	101	544	612	653
Ürün Şablonu 17	112	146	696	748	856
Ürün Şablonu 18	50	64	392	568	712
Ürün Şablonu 19	158	312	1098	1156	1378
Ürün Şablonu 20	78	197	499	653	834

3.1.2.3 ORB

ORB [14] özellik çıkarma yöntemi, FAST [12] ve BRIEF [13] yöntemlerinin bazı modifikasyonlarla birleşimidir. İlk olarak FAST, görüntünün ölçek uzay piramidindeki önemli noktalarını tespit eder. Daha sonra anahtar noktaları sıralamak ve bunların en üst N âdetini bulmak için Harris köşe ölçüsü uygulanır. Daha sonra, köşesi merkezde bulunan yamanın yoğunluk ağırlıklı merkezi boyunca yerel yönelimin bir dönme matrisi hesaplanır. Bundan sonra BRIEF tanımlayıcıları oryantasyona göre yönlendirilir ve ikili dizi ORB tanımlayıcısı [14] olarak tutulur. Son olarak eşleştirme adımında BRISK'deki [11] gibi Hamming mesafesi kullanılır ve elde edilen değer eşik değerden küçükse görsel tanımlayıcıları eşleştirilir.

3.1.2.4 SURF, BRISK ve ORB'nin Hibrit Kullanımı

Her özellik çıkarma yöntemi için, ürün şablonu ile bölge teklifi arasındaki eşleşen özellik sayısının ürün şablonunda bulunan özellik sayısına oranı olarak hesaplanan eşleştirme puanı, belli bir eşik değerinden büyükse, bu bölge teklifinin bu eşleştirme puanıyla bu ürüne sınıflandırıldığı sonucu çıkarılır. Bölge önerilerini ve ürün şablonlarını eşleştirmek için ayrı ayrı kullanılan SURF [10], BRISK [11] ve ORB [14] özellik çıkarma yöntemlerine ek olarak, farklı ürünler için farklı özelliklerin daha ayırt edilebilir olmasından dolayı bu üç yöntemin hibrit kullanımı önerilmiştir. Bunun için, her bir ürün sınıfına ait tüm bölge önerileri için üç farklı özellik çıkarım yöntemi ile elde edilen eşleştirme puanları, sınıf bazında normalize edilerek birbirleriyle karşılaştırılmıştır. Daha sonra en yüksek eşleştirme puanına sahip özellik çıkarım yöntemi, o ürün sınıfını tanımaya özgü yöntem olarak seçilmiştir.

3.1.3 İyileştirme Aşaması

Ürün sınıflandırma aşamasından, her raf görseli için ürün sınıfı bilgilerinin ve eşleştirme puanının yer aldığı sınırlayıcı kutular elde edilir. Raf görsellerinde yer alan her ürün için birden fazla üst üste binen kutuların en doğru şekilde tek bir sınırlayıcı kutuya indirgenecek şekilde birleştirilmesi gerekmektedir. Bu amaçla, iyileştirme aşamasında, ürün sınıfları ve konum bilgileri üzerinde [15]'teki kümeleme algoritmasına (CAI, Clustering Algorithm) benzer ancak bazı değişikliklerle iyileştirmeler yapılan bir algoritma ile nihai sonuca ulaşılmıştır. Önerilen yöntemin [15]'ten temel farkı, kümeleme algoritmasına komşuluk ilişkisinin eklenmesidir. Market raflarında aynı ürünlerin yan yana konulma

olasılığının yüksek olduğu bilgisinden yararlanılarak, kümeleme algoritmasına komşuluk ilişkisinin eklenmesinin çalışmaya olumlu bir katkı sağlaması beklenmektedir. Komşulukla ilgili kümeleme algoritması (NR-CAlg, Neighborhood-Related Clustering Algorithm) aşağıdaki adımlardan oluşur:

1. Her bir test görüntüsü için elde edilen bölge önerileri en yüksek puandan en düşük puana doğru sıralanır.
2. En yüksek puanı alan bölge önerisi ilk kümenin ilk elemanı olarak alınır.
3. Sıralamadaki bölge önerisi ile oluşturulan kümenin kesişim alanı, kendi alanının yarısından büyükse bu bölge önerisi o kümeye dâhil edilir. Bölge teklifi, mevcut kümelerden herhangi birine atanamıyorsa o bölge teklifi ile yeni bir küme oluşturulur.
4. Bir küme, aynı kümedeki her sınırlayıcı kutu değerinin ortalaması, eşleşen puanın maksimumu ve sınıf bilgisiyle temsil edilir.
5. Bu işlemler kümelenmemiş bölge teklifi kalmayana kadar devam eder.
6. Elde edilen tüm kümelere komşuluk ilişkisi eklemek için küme çiftlerinin birbirine en yakın iki noktası arasındaki uzaklık matrisi hesaplanır.
7. Bu mesafenin bir eşik değerinden (her kümenin sınırlayıcı kutu bilgisinde yer alan genişlik değerinin yarısı kadar) az olduğu durumlarda, ürünlerin yan yana olduğu varsayılır ve eşleştirme puanlarının değerleri kendi puanlarının 1/10'u oranında artırılır.
8. Yeni puan durumuna göre, tüm kümelerden alınan maksimum puanın %40'ından az eşleştirme puanına sahip olan kümeler elenir.

Bu sürecin sonunda ürün sınıfları ve lokalizasyonlarına ilişkin nihai bilgiler elde edilir.

NR-CAlg algoritması, farklı raf tasarımları, farklı ürün boyutları, ürün çeşidinin çok olması, sahne yoğunluğunun fazla olması gibi durumlarla başa çıkabilmektedir. Kullanılan parametrelerin her raf görseli için özel olarak hesaplanması, NR-CAlg'nin farklı durumlarda da aynı performansı göstermesini sağlamaktadır. NR-CAlg algoritmasında üç farklı parametre kullanılmaktadır: (1) bölge önerisi ile herhangi bir kümenin minimum kesişim alanı, (2) komşuluk ilişkisinin varlığını kabul etmek için minimum uzaklık ölçüsü ve (3) elenecek kümelere karar vermek için kullanılan minimum puan. Parametre (1) için kullanılan eşik değeri %50'dir. Karşılaştırılan bölge tekliflerinin örtüşen alanı %50'den fazla olması bu bölge

teklifinin en çok o ürüne ait olduğunu gösterir ve o ürün kümesine ait kümeye dâhil edilir. Parametre (2) için kullanılan eşik değeri, her kümenin sınırlayıcı kutu bilgisinde yer alan genişlik değerinin yarısı olarak belirtilir. Bu değer, ürünlerin birbirine yakın olduğunu ve aralarında başka ürün bulunmadığını ifade eder. Eleme işlemi için kullanılan eşik (2) değeri, tüm kümelerden alınacak maksimum puanın %40'ı olarak belirlenir.

3.2 Kullanılan Veri Setleri

Bu çalışmanın eğitim ve test aşamaları için Bölüm 2’de incelenen veri setleri içinden en kapsamlı, ince taneli ve yoğun ürünlerin bulunduğu farklı özelliklerdeki veri setleri seçilmiştir. Kullanılan veri setleri Grocery Products [6] ve alt kümeleri GP-20 [15] ile GP-181 [16], Grocery Dataset [17] ve SKU-110K'nın [5] alt kümesine ait örnek görseller Şekil 3.8’de, veri setlerine ait detaylı bilgiler ise Tablo 3.2’de verilmiştir. Önerilen yaklaşımın ilk aşaması için raftaki tüm ürünlerin etiketlendiği bir veri seti gerekmektedir. ACF [7], YOLOv4 [8] ve SSD’nin [9] eğitimi için Grocery Dataset [17] ve SKU-110K'nın [5] bir alt kümesi birleştirilerek yeni bir veri seti oluşturulmuştur. Önerilen uçtan-uca sistemin performansını değerlendirmek ve karşılaştırılabilir sonuçlar sağlamak için, farklı sayıda ürün çeşidine ve test görsellerine sahip mevcut Grocery Products [6] ve alt kümeleri GP-20 [15], GP-181 [16]) kullanılmıştır.

Önerilen yaklaşımın ilk aşaması ürün konumlarının tespiti için Grocery Dataset [17] ve SKU-110K [5], raf görsellerindeki ürünlerin sayısı ve yoğunluğu bakımından farklılık göstermektedir. Grocery Dataset [17] sadece aynı tip sigara paketi görüntülerinden oluşurken, SKU-110K [5] veri seti ürün çeşitliğinin fazla olduğu yoğun raf görüntülerinden oluşur. Bu nedenle, ilk aşamada detektörleri eğitmek için bu iki veri kümesinden alınan örnekler bir araya getirilerek eğitim kümesindeki çeşitlilik artırılmıştır. Bu amaçla SKU-110K [5] veri setinden doğrulama için ayrılmış olan 588 raf görseli ile Grocery Dataset [17] test kümesindeki 354 raf görselinin tamamı birleştirilmiştir (GD + SKU-110K_{val}). Raf görsellerinde yer alan tüm ürünler Şekil 3.8’de gösterildiği gibi etiketlidir.

Tablo 3.1 Kullanılan veri setlerinin özellikleri

Kullanılan Aşama	Veri Seti	Ürün Çeşidi	Görüntü Sayısı	Açıklama Dosyası	Açıklamalı Ürünler
Aşama-1'in eğitimi için	Grocery Dataset	1	354 raf görüntüsü	ürüne özgü sınırlayıcı kutu	tüm ürünler
	SKU-110Kval	1	588 raf görüntüsü	ürüne özgü sınırlayıcı kutu	tüm ürünler
Aşama-2'nin eğitimi için	GP-20	20	Her ürün için 1 şablon görüntü	-	-
	GP-181	181	Her ürün için 1 şablon görüntü	-	-
	Grocery Products	27 üst-sınıf	3.235 şablon görüntü (ürün başına 25-415 görsel)	-	-
		3235 alt-sınıf	Her ürün için 1 şablon görüntü	-	-
Uçtan-uça sistemin testi için	GP-20	20	71 raf görüntüsü	ürüne özgü sınırlayıcı kutu	Seçili ürünler
	GP-181	181	73 raf görüntüsü	ürüne özgü sınırlayıcı kutu	Tüm ürünler
	Grocery Products	27 üst-sınıf	680 raf görüntüsü	Aynı ürün grubu için tek sınırlayıcı kutu	Seçili ürünler
		3235 alt-sınıf	680 raf görüntüsü	Aynı ürün grubu için tek sınırlayıcı kutu	Seçili ürünler



Şekil 3.1 Kullanılan veri setlerinden örnek görüntüler: (a,b) SKU-110K'nın örnek raf görüntüleri [5]; (c,d) Grocery Dataset'in örnek raf görüntüleri [17]; (e,f) GP-20'nin örnek raf görüntüleri [15]; (g,h) GP-181'in örnek raf görüntüleri [16]; (i,j) Grocery Products'ın örnek raf görselleri [6]; (k,l) Grocery Products ürünlerinin örnek şablon görüntüleri [6]

Çok aşamalı uçtan-uca ürün tanıma sisteminin performansı test etmek için kullanılan veri setleri Grocery Products [6], GP-20 [15] ve GP-181 [16] Tablo 3.2'de de gösterildiği üzere, ürün çeşidi, görüntü sayısı ve etiketlenen ürünler bakımından farklılık göstermektedir. Grocery Products [6] veri seti, market raflarında ürün tanıma problemi için hem ürün çeşitliliğinin ve görüntü sayısının fazla olması hem de ince taneli yapısı nedeniyle literatürde yaygın olarak kullanılan koleksiyonlardan biridir. Ayrıca, Grocery Products [6] ve GP-20'deki [15] raf görselleri sadece seçili ürünler için açıklamalı sınırlayıcı kutu bilgisine sahipken, GP-181 [16] veri setinde rafta yer alan tüm ürünler için açıklamalı sınırlayıcı kutu bilgisine sahiptir. Bu nedenlerle, önerilen yaklaşımın performansını mevcut

alıřmalarla karřılařtırmak iin Grocery Products [6], GP-20 [15] ve GP-181 [16] veri kmeleri seilmiřtir. Mevcut alıřmaların [6], [15], [16], [36], [47] ve [50] kapsadıėı eėitim ve test rnekleri bu alıřmada da deėiřtirilmeden Tablo 3.2’de gsterilen řekilde kullanılmıřtır. ncelikle, bu alıřmada farklı zellik ıkarma yntemlerini karřılařtırmak iin en kk lekli veri seti olan GP-20 [15] kullanılmıřtır. Daha sonra GP-20’de [15] en yksek performansı gsteren nerilen yaklařım GP-181 [16] ve Grocery Products [6] veri setleri ile test edilmiřtir.



4.1 GP-20 Veri Setinde Alınan Sonuçlar

Veri setleri arasında en küçük ölçekli olan GP-20 veri seti [15] üzerinde, Şekil 3.1’de gösterilen çok aşamalı uçtan-uca ürün tanıma sisteminin her aşaması için önerilen yöntemler test edilmiştir. Aşama-1 için geleneksel bilgisayarla görme yaklaşımı, ACF, YOLOv4 ve SSD, aşama-2 için SURF, BRISK, ORB ve bunların hibrit kullanımı, aşama-3 için de kümeleme algoritması ve komşuluk ilişkili kümeleme algoritması kullanılarak deneysel sonuçlar elde edilmiştir.

Çok aşamalı uçtan-uca ürün tanıma sisteminin her aşamasının performansı hem ayrı ayrı, hem de bütün bir sistem performansı olarak incelenmiştir. Bunun için öncelikle aşama-1, aşama-2 ve aşama-3 için önerilen yöntemlerin, sonrasında da kullanılan tüm yöntemlerin uçtan-uca performansı GP-20 veri seti üzerinde alınan sonuçlarla verilmiştir.

4.1.1 Aşama-1 için Önerilen Yöntemlerin Performansı

Çalışmanın 1. aşaması olan ürün konumlarının tespiti aşamasında kullanılan yöntemler geleneksel bilgisayarla görme yaklaşımı, ACF, YOLOv4 ve SSD’nin GP-20 veri setinde yer alan 71 raf görüntüsü üzerindeki performansı Tablo 4.1’de verilmiştir. Önerilen geleneksel bilgisayarla görme yaklaşımı herhangi bir eğitim gerektirmezken, kullanılan diğer yöntemler ACF, YOLOv4 ve SSD’nin eğitimi için Grocery Dataset ve SKU-110K veri setlerinden oluşturduğumuz eğitim seti (GD+SKU-110Kval) kullanılmıştır.

Yöntemlerin performansı, GP-20 veri setinin oluşturulduğu çalışma [15]’te kullanılan ayrık metriğe göre eşitsizlik (4.1) ve (4.2)’de gösterilen değerlendirme kriterlerine göre hesaplanmıştır. Çalışma [15]’te tanımlandığı üzere, ürün için tespit edilen sınırlayıcı kutunun orta noktası, ürünün temel doğruluk kutusunun içinde yer alıyorsa, ürün gerçek pozitif (TP, True Positive) olarak; aksi takdirde yanlış pozitif (FP, False Positive) olarak kabul edilmiştir.

Ayrık metrik kesinlik değeri, pozitif olarak tahmin edilen ürünlerin gerçekten kaç tanesinin pozitif olduğunu gösterir. Bunun için, doğru olarak tespit edilmiş

ürünlerin (TP) toplam tespit edilen ürün sayısına (TP+FP) oranı hesaplanır [15]. Ayırık metrik duyarlılık değeri, pozitif olarak tahmin etmemiz gereken ürünlerin ne kadarının pozitif olarak tahmin edildiğini gösterir. Bunun için, doğru olarak tespit edilen ürünlerin (TP) etiketli ürün sayısına (TP+FN (yanlış negatif, False Negative) oranı hesaplanır [15].

$$Kesinlik = \frac{TP}{TP + FP} \quad (4.1)$$

$$Duyarlılık = \frac{TP}{TP + FN} \quad (4.2)$$

Tablo 4.1.'de alınan sonuçlara baktığımızda ürün konumlarının tespiti için önerilen yöntemlerden en iyi performans SSD kullanılarak elde edilmiştir. Performans sıralaması yine DNN tabanlı bir detektör olan YOLOv4, geleneksel bilgisayarla görme yaklaşımı ve öğrenmeye dayalı ACF detektörü ile devam etmektedir.

Tablo 4.1 Aşama-1 için önerilen yöntemlerin GP-20 veri setindeki sonuçları

Yöntemler	Kesinlik	Duyarlılık
Geleneksel Bilgisayarla Görme Yaklaşımı	84,4	87,2
ACF	82,7	78,3
YOLOv4	89,6	91,5
SSD	<u>91,3</u>	<u>94,2</u>

4.1.2 Aşama-2 için Önerilen Yöntemlerin Performansı

Çalışmanın 2. aşaması olan ürün sınıflandırma aşamasında kullanılacak özelliklerin seçimi için GP-20 veri seti kullanılmıştır. Bölge önerilerini ve ürün şablonlarını eşleştirmek için ayrı ayrı kullanılan SURF, BRISK ve ORB özellik çıkarma yöntemlerine ek olarak, farklı ürünler için farklı özelliklerin daha ayırt edilebilir olmasından dolayı bu üç yöntemin hibrit kullanımı önerilmiştir. Bu noktada sistemin başarısını aşama-1 ve aşama-3'ten bağımsız değerlendirmek için potansiyel bölge önerileri GP-20 veri setindeki raf görüntülerinden kesilerek elde

edilmiştir. Sonrasında GP-20 veri setindeki her bir ürün sınıfı için oluşturduğumuz bölge önerileri üzerinden üç farklı özellik çıkarım yöntemiyle özellik eşleştirmesi yapılmıştır.

Tablo 4.2’de görüldüğü üzere SURF, BRISK ve ORB özellikleri ile GP-20 veri setindeki her ürün sınıfı için elde edilen eşleştirme puanları sınıf bazında normalize edilerek ortalamaları alınmıştır. Her özellik çıkarma yöntemi ile elde edilen ürün tanıma sonuçları da tabloda yer almaktadır. Verilen sonuçlardaki gösterim şeklinde (.../.../...) ilk sayı sınıflandırılan ürün sayısını, ikincisi sınıflandırılması gereken ürün sayısını, üçüncüsü de doğru olarak sınıflandırılan ürün sayısını göstermektedir. SURF, BRISK ve ORB özellikleri tek başına kullanıldığında toplamda kaç ürün tespit edildiği ve tespit edilmesi gereken 190 üründen kaçının doğru olarak tespit edilebildiği Tablo 4.2’nin en altındaki satırda verilmiştir.

Elde edilen özellik skoru ortalamalarından vurgulanan değerler o ürün sınıfı için elde edilen en yüksek değeri ve tanıma sonucunu göstermektedir. En yüksek skoru alan özellik 19. ürün sınıfı haricinde tanıma sonucu olarak da en iyi performansı göstermiştir. Elde edilen sonuçlardan, farklı sınıflar için farklı özelliklerin daha iyi sonuç verdiğini görmek mümkündür. 5 ürün sınıfı için BRISK özelliği, 7 ürün için ORB özelliği ve 8 ürün sınıfı için SURF özelliği daha iyi sonuç vermiştir. Bu durum özelliklerin birlikte kullanımının, sistemin farklı öznetelik çıkarma yöntemlerinin güçlü yönlerinden yararlanmasına olanak tanıyarak başarıyı arttıracaklarını göstermektedir. Bu nedenle, en yüksek eşleştirme puanına sahip özellik çıkarım yöntemi, o ürün sınıfını tanımaya özgü yöntem olarak seçilmiştir.

Ayrıca, Tablo 4.3’te SURF, BRISK ve ORB’nin ayrı ayrı ve hepsinin birlikte kullanım durumlarındaki sınıflandırılan toplam ürün sayısı ve doğru olarak sınıflandırılan ürün sayıları verilmiştir. GP-20 veri setindeki 71 raf görüntüsünde tespit edilmesi gereken 190 ürün bulunmaktadır. Alınan sonuçlara bakıldığında tespit edilmesi gereken ürünlerden en fazla doğru sınıflandırılan ürün sayısına SURF, BRISK ve ORB’nin hibrit kullanımı ile erişildiği gözükmemektedir. Ayrıca, tespit edilen ürün sayısı ile doğru tespit edilen ürün sayısı arasındaki farkın en az olduğu değere de bu üç özelliğin bir arada kullanılması ile ulaşılmıştır. Bu durum, TP sayısını artırırken FP sayısını azaltarak sistemin başarısının arttığını göstermektedir.

Tablo 4.1 SURF, BRISK ve ORB özelliklerinin GP-20 veri setindeki sonuçları

Ürün Sınıfı	SURF Ort.	SURF Sonucu	BRISK Ort.	BRISK Sonucu	ORB Ort.	ORB Sonucu
1	0,1286	4/6/4	0,1533	5/6/4	<u>0,1896</u>	<u>5/6/5</u>
2	0,0978	15/17/10	0,1588	10/17/10	<u>0,1657</u>	<u>13/17/12</u>
3	0,1981	3/4/2	<u>0,2602</u>	<u>3/4/3</u>	0,2167	6/4/3
4	<u>0,2189</u>	<u>1/1/1</u>	0,0942	2/1/1	0,1099	1/1/1
5	0,1056	6/10/4	0,1095	7/10/5	<u>0,1567</u>	<u>11/10/9</u>
6	<u>0,3546</u>	<u>25/25/21</u>	0,2097	14/25/12	0,3349	17/25/15
7	0,1679	10/19/9	<u>0,2181</u>	<u>14/19/12</u>	0,1790	14/19/8
8	0,1456	5/9/4	0,1636	8/9/7	<u>0,2564</u>	<u>9/9/8</u>
9	<u>0,2256</u>	<u>6/6/5</u>	0,1129	9/6/6	0,0587	15/6/5
10	<u>0,1685</u>	<u>3/3/3</u>	0,2507	3/3/3	0,1186	3/3/3
11	0,1451	8/11/7	<u>0,2814</u>	<u>11/11/11</u>	0,2641	10/11/10
12	0,1202	6/10/5	<u>0,2290</u>	<u>8/10/7</u>	0,1834	6/10/6
13	<u>0,3201</u>	<u>18/18/15</u>	0,2529	22/18/12	0,1525	22/18/10
14	0,1453	8/9/5	0,1747	6/9/6	<u>0,3029</u>	<u>7/9/7</u>
15	<u>0,1651</u>	<u>8/8/7</u>	0,0830	5/8/5	0,0408	5/8/4
16	0,1298	4/4/3	<u>0,1922</u>	<u>4/4/4</u>	0,1000	4/4/3
17	0,1967	7/6/5	0,1530	5/6/4	<u>0,2298</u>	<u>6/6/6</u>
18	0,1741	8/5/4	0,1831	8/5/5	<u>0,2472</u>	<u>7/5/5</u>
19	0,1256	11/12/5	<u>0,1445</u>	<u>15/12/10</u>	0,1118	15/12/11
20	<u>0,2152</u>	<u>9/7/6</u>	0,1134	5/7/4	0,1603	13/7/6
Top.		165/190/125		164/190/131		189/190/137

Tablo 4.2 Aşama-2 için önerilen yöntemlerin GP-20 veri setindeki sonuçları

Yöntemler	Tespit Edilen Ürün Sayısı	Doğru Tespit Edilen Ürün Sayısı
SURF	165	125
BRISK	164	131
ORB	189	137
SURF&BRISK&ORB	<u>182</u>	<u>157</u>

4.1.3 Aşama-3 için Önerilen Yöntemlerin Performansı

Çalışmanın son aşaması olan iyileştirme aşamasında, kümeleme algoritmasına eklenen komşuluk ilişkisinin katkısını göstermek amacıyla CAlg ile NR-CAlg karşılaştırılmıştır. GP-20 veri seti üzerinde NR-CAlg ile CAlg'nin performansı Tablo 4.4'te gösterilmektedir. Çalışmanın bu aşamasını aşama-1 ve aşama-2'den bağımsız değerlendirmek mümkün değildir. Bu nedenle, önceki aşamalarda en iyi performansı gösteren aşama-1 için SSD ve aşama-2 için SURF, BRISK ve ORB'nin hibrit hali iyileştirme aşamasının performans karşılaştırılması için kullanılmıştır. Elde edilen kesinlik ve duyarlılık değerleri, kümeleme algoritmasına komşuluk ilişkisinin eklenmesinin önerilen yöntemin performansına olumlu katkı sağladığını göstermektedir. Kesinlik değeri 0,7, duyarlılık değeri ise 3,8 artmıştır.

Tablo 4.1 Aşama-3 için önerilen yöntemlerin GP-20 veri setindeki sonuçları

	Kesinlik	Duyarlılık
CAlg	78,1	77,5
NR-CAlg	<u>78,8</u>	<u>81,3</u>

4.1.4 Çok Aşamalı Uçtan-uca Ürün Tanıma Yaklaşımı için Kullanılan Tüm Yöntemlerin Performansı

Aşama-1 ve aşama-2 için kullanılan tüm yöntemler birlikte ele alınarak GP-20 veri seti üzerinde uçtan-uca ürün tanıma sisteminin performansı ölçülmüştür. Sistemin uçtan-uca ele alınabilmesi için aşama-3'te önerilen komşuluk ilişkili kümeleme algoritması kullanılmıştır. Önerilen yöntemlerin GP-20 veri seti üzerindeki

performansı, 4.1.1’de kullanılan, eşitsizlik (4.1) ve (4.2)’de gösterilen kesinlik ve duyarlılık değerlerine göre hesaplanmıştır.

Önerilen çok aşamalı uçtan-uca ürün tanıma yaklaşımının GP-20 veri setinde yer alan 71 raf görüntüsü üzerinde 20 farklı market ürününü tanıma performansı Tablo 4.5’te gösterilmektedir. Tablonun ikinci satırında ürün tanıma sisteminin ilk adımı için kullanılan farklı yöntemler listelenmiştir. Tablonun ikinci sütununda ise ürün tanıma sisteminin ikinci aşamasında kullanılan farklı yöntemler listelenmiştir. Tabloda yer alan her kesinlik ve duyarlılık değeri, ilgili olduğu satır ve sütunda temsil edilen yöntemlerden sonra komşuluk ilişkili eklenmiş kümeleme algoritması uygulandıktan sonra bulunmuştur.

Tablo 4.1 Çok aşamalı uçtan-uca ürün tanıma yaklaşımı için kullanılan tüm yöntemlerin GP-20 veri setindeki sonuçları

		Ürün Konumunun Tespiti için Kullanılan Yöntemler							
		Geleneksel Bilgisayarla Görme Yaklaşımı		ACF*		YOLOv4*		SSD*	
		Kesinlik	Duyarlılık	Kesinlik	Duyarlılık	Kesinlik	Duyarlılık	Kesinlik	Duyarlılık
Ürün Sınıflandırması için Kullanılan Yöntemler	SURF	61,6	<u>82,2</u>	59,2	<u>72,3</u>	67,5	75,2	71,5	78,2
	BRISK	67,6	65,2	57,7	66,8	73,1	78,2	73,7	<u>83,2</u>
	ORB	<u>76,5</u>	58,1	48,5	68,9	75,3	69,6	78,1	71,6
	SURF & BRISK & ORB	75,1	78,1	<u>62,7</u>	71,2	<u>77,2</u>	<u>79,5</u>	<u>78,8</u>	81,3

* Detektörün eğitimi için GD + SKU-110Kval veri seti kullanılmıştır.

Tablo 4.5'teki sonuçlardan görüleceği üzere en yüksek kesinlik ve duyarlılık değerleri farklı yöntemlerle elde edilmiştir. Kullanılan DNN tabanlı modeller YOLOv4 ve SSD diğer modellere göre daha yüksek performans göstermiştir. Ayrıca, elde edilen kesinlik ve duyarlılık değerleri çiftleri arasındaki optimum sonuca aşama-1'de SSD ve aşama-2'de SURF, BRISK ve ORB'nin hibrit kullanımıyla ulaşılmıştır.

Tablo 4.6'da ürün tanıma probleminin her bir aşamasında en iyi performansı gösteren yöntemlerin (SSD + SURF&BRISK&ORB + NR-CAlg) GP-20 veri seti üzerindeki karışıklık matrisi (confusion matrix) analizi verilmiştir. Normal bir sınıflandırma probleminin karışıklık matrisinden farklı olarak nesne tespiti için oluşturulan karışıklık matrisinde tahminlenen sınıfların yanında “tespit edilmeyen” ve temel-gerçek sınıfların yanında “hayalet tahmin” ifadeleri yer alır [103].

Tablo 4.2 Önerilen yöntemin (SSD + SURF&BRISK&ORB + NR-CAlg) GP-20 veri seti üzerindeki karışıklık matrisi analizi

Temel-gerçek																					
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Hayalet tahmin
Tahminlenen	1	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
	2	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
	3	0	1	3	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	4	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
	5	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
	6	0	0	0	0	0	21	0	0	1	0	0	1	0	0	0	0	0	0	0	4
	7	0	0	0	0	0	0	12	0	0	0	0	0	0	0	0	0	0	0	0	2
	8	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0	1	3
	9	0	0	0	0	0	1	0	0	5	0	0	0	0	0	0	0	0	0	0	0
	10	0	0	0	0	0	0	1	1	0	3	0	0	0	0	0	0	0	0	0	0
	11	0	0	0	0	0	0	0	0	0	0	10	1	1	0	0	0	0	0	0	0
	12	0	0	0	0	0	0	0	0	0	0	0	7	1	0	0	0	0	0	0	0
	13	0	0	0	0	0	0	1	0	0	0	0	0	15	0	0	0	0	0	0	3
	14	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0
	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0
	16	0	0	0	0	1	0	0	0	0	0	0	0	0	0	4	0	1	1	0	0
	17	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	6	0	0	0	0
	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	1	0	0
	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	2
	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	1
	Tespit edilmeyen	1	4	1	0	0	1	5	0	0	0	1	0	0	2	0	0	0	0	0	

Tespit edilmeyen ifadesi temel-gerçek sınıf bilgisine sahip olduğumuz ürünün herhangi bir tahminimizle eşleşmediği durumlar için kullanılır. Örnek olarak, 1. ürün sınıfı için baktığımızda 5 ürünün doğru olarak tespit edildiği fakat 1 ürünün herhangi bir sınırlayıcı kutu ve sınıf tahminimizle eşleşmediği (tespit edilemeyen) görülmektedir. Hayalet tahmin ise temel-gerçek etiketlerde mevcut olmayan ürünler tespit edilmesi anlamına gelir. Örnek olarak, 1. ürün sınıfı için baktığımızda doğru tespit edilen 5 ürün ve tespit edilemeyen 1 ürünün yanı sıra bu temel-gerçek bilgilerde mevcut olmayan 2 ürün tespitinin daha var olduğu görülmektedir. Tabloda yeşil ile gösterilen değerler TP sonuçları, kırmızı ile gösterilen değerler de FP sonuçları göstermektedir.

Bu sonuçların yanı sıra, bu tez çalışmasında önerilen uçtan-uca ürün tanıma sisteminin performansını kanıtlamak için GP-20 veri setinin kullanıldığı başka bir çalışma [15] ile karşılaştırma yapılmıştır. Karşılaştırma için elde edilen kesinlik ve duyarlılık değerleri ile Tablo 4.7’de verilmiştir. Bu tez çalışmasında önerilen yöntem Tablo 4.7’de SSD + SURF&BRISK&ORB + NR-CAlg olarak gösterilmiştir. Ek olarak Tablo 4.7’de, çalışma [15]’te kullanılan iki farklı yöntemle elde edilen sonuçlar: (1) önerilen DNN (AlexNet) ve (2) önerilen BoW (SURF) olarak verilmiştir. Önerilen DNN yönteminde AlexNet ağı kullanılarak, önerilen BoW yönteminde ise SURF özellikleri kullanılarak ürün şablonları ile bölge önerileri arasında özellik eşleştirmesi yapılmıştır. Bu tez çalışmasında önerilen yöntemle elde edilen 78,8 kesinlik ve 81,3 duyarlılık değerleri ile çalışma [15]’te önerilen her iki yöntemden de daha yüksek bir performans gösterilmiştir.

Tablo 4.3 Önerilen yöntemin (SSD + SURF&BRISK&ORB + NR-CAlg) GP-20 veri seti üzerindeki performansının yapılan çalışmalarla karşılaştırılması

	Kesinlik	Duyarlılık
[15]’te önerilen DNN (AlexNet)	73,1	73,6
[15]’te önerilen BoW (SURF)	77,7	76,5
SSD + SURF&BRISK&ORB +NR-CAlg	<u>78,8</u>	<u>81,3</u>

Tüm bu sonuçlara bakıldığında, ürün tespiti için DNN tabanlı SSD kullanılmasının, farklı özellik çıkarım yöntemlerinin hibrit olarak kullanılmasının ve kümeleme algoritmasına komşuluk ilişkisinin eklenmesinin sistemin performansına önemli katkı sağladığını söylemek mümkündür. Böylece, diğer veri setlerinin performans değerlendirmesi için en uygun uçtan-uca ürün tanıma yöntemi (SSD + SURF&BRISK&ORB + NR-CAlg) elde edilmiştir.

4.2 GP-181 Veri Setinde Alınan Sonuçlar

Bu tez çalışmasında önerilen uçtan-uca yöntemin GP-181 veri seti [16] üzerindeki performansı, [47]'de tanımlanan mAP ve ürün duyarlılığına (PR, Product Recall) göre hesaplanmıştır. Bir ürünün tahmin edilen ve temel gerçek sınırlayıcı kutularının kesiştirilmiş bölgeleri (IoU, Intersection over Union) 0,5'ten büyükse o ürün TP olarak kabul edilmiştir; aksi takdirde FP olarak kabul edilmiştir. IoU değeri, tahmin edilen ve temel gerçek sınırlayıcı kutunun kesiştikleri alanın, bileşimleri alanına bölünmesi ile hesaplanır. Çalışma sonucundaki mAP değeri, detektörün kesinlik-duyarlılık eğrisinin altında kalan alanın yaklaşık değeri olarak ölçülmüştür. PR değeri ise tüm raf görüntülerindeki ortalama ürün duyarlılığı olarak ölçülmüştür [47].

Daha önce yayınlanmış çalışmalarla [16], [47], elde edilen karşılaştırmalı sonuçlar Tablo 4.8'de verilmiştir. Çalışma [16]'da, bu tez çalışmasından farklı olarak ürün konumlarını belirlemek ve kaçırılan ürünleri tespit etmek için planogram bilgileri kullanılmış olsa da, önerilen uçtan-uca ürün tanıma sistemi SSD + SURF&BRISK&ORB + NR-CAlg, [16]'dan daha yüksek mAP ve PR değerleri elde etmiştir.

Ek olarak, önerilen yöntem ile elde edilen PR değeri [47]'ye yaklaştığında elde edilen mAP değeri [47]'den ~%5 daha yüksektir. Önerilen yöntem ile GP-20 veri setine göre GP-181 veri setinde daha yüksek sonuçlar elde edilmesinin nedeni, GP-20'deki raf görüntülerinde sadece önceden belirlenmiş 20 ürün için ürün tanıma işleminin yapılması, GP-181 veri setindeyse raf görüntülerinde yer alan tüm ürünlerin etiketlenmiş olmasıdır. Böylece, raf görsellerindeki tüm ürünlerin

tanınmasıyla çalışmadaki FP sayısının azalması sistemin performansının artmasını sağlamıştır. Ürün tespiti için çalışma [47]'de kullanılan YOLOv2 yerine bu çalışmada SSD'nin kullanılması, YOLO'nun görüntülerdeki daha küçük nesneleri tanımlanmasındaki sınırlamasının ortadan kaldırılmasına da katkıda bulunmuştur.

Tablo 4.4 Önerilen yöntemin (SSD + SURF&BRISK&ORB + NR-CAlg) GP-181 veri seti üzerindeki performansının yapılan çalışmalarla karşılaştırılması

	mAP	PR
[16]	66,37	75,0
[47]	76,93	85,71
SSD + SURF&BRISK&ORB +NR-CAlg	81,23	84,57

4.3 Grocery Products Veri Setinde Alınan Sonuçlar

Önerilen çok aşamalı uçtan-uca ürün tanıma yönteminin Grocery Products veri seti [6] üzerindeki performansı, [36]'da tanımlanan kategorileştirme doğruluğu (CA, Categorization Accuracy), ürün doğruluğu (PA, Product Accuracy), ürün kesinliği (PP, Product Precision) ve PR metriklerine göre hesaplanmıştır. CA metriği, eşitsizlik (4.3)'te olduğu gibi, 27 gıda kategorisinin her bir sınıfı için doğru sınıflandırılmış görüntü sayısının toplam görüntü sayısına oranlarının ortalaması olarak hesaplanmıştır. PA metriği, eşitsizlik (4.4)'te görüldüğü üzere 3235 ince taneli sınıfın her biri için temel-gerçek ve tahmin edilen sınırlayıcı kutuların kesişiminin, bunların birleşimine oranlarının ortalaması olarak hesaplanmıştır. PP metriği ise, eşitsizlik (4.5)'te gösterildiği gibi, 3235 ince taneli sınıfının her biri için temel gerçek ve tahmin edilen sınırlayıcı kutuların kesişiminin, ürünün tahmin edilen sınırlayıcı kutusuna oranlarının ortalaması olarak hesaplanmıştır. PR metriği de, eşitsizlik (4.6)'da olduğu gibi, 3235 ince taneli sınıfın her biri için temel-gerçek ve tahmin edilen sınırlayıcı kutuların kesişiminin, ürünün temel-gerçek kutusuna oranlarının ortalaması olarak hesaplanmıştır. (4.3)'ten (4.6)'ya kadar tüm eşitsizliklerde yer alan T , toplam sınıf sayısını; k_l , l sınıfında doğru olarak sınıflandırılan görsellerin sayısını; n_l , l sınıfındaki toplam görsel sayısını, C_l ,

etiketli ürünün temel-gerçek sınırlayıcı kutusunu ve C'_l ürün için tahmin edilen sınırlayıcı kutuyu göstermektedir.

$$CA = \frac{1}{T} \sum_{l=1}^T \frac{k_l}{n_l} \quad (4.3)$$

$$PA = \frac{1}{T} \sum_{l=1}^T \frac{C_l \cap C'_l}{C_l \cup C'_l} \quad (4.4)$$

$$PP = \frac{1}{T} \sum_{l=1}^T \frac{C_l \cap C'_l}{C'_l} \quad (4.5)$$

$$PR = \frac{1}{T} \sum_{l=1}^T \frac{C_l \cap C'_l}{C_l} \quad (4.6)$$

Daha önce yayınlanmış çalışmalarda elde edilen sonuçlarla [6], [36] ve [50], bu tez çalışmasında elde edilen sonuçlar karşılaştırmalı olarak Tablo 4.9'da verilmiştir. Ürün tespiti ve sınıflandırması aşamalarının ortak olarak ele alındığı çalışma [6]'da ki ürün tanıma süreci, yoğun SIFT özelliklerinin eşleştirilmesi ve rastgele orman algoritması kullanılarak çoklu sınıf sıralaması, sonrasında da genetik algoritma kullanılarak optimizasyon eşleştirmesi yapılarak tamamlanır. Çalışma [36]'da market ürünlerini lokalize etmek ve sınıflandırmak için SURF özellikleri ile Hough dönüşümü kullanılmıştır. Diğer çalışmalardan farklı olarak [50]'de ise, OCR teknikleri kullanılarak ürün paketleri üzerindeki kelimeler tespit edilerek ürünler tanımlamaya çalışılmış; daha sonra ayırt edici yamalar kullanılarak ürünlerin görsel özellikleri çıkartılmış ve SVM kullanılarak ürünler sınıflandırılmıştır. Bu tez çalışmasında önerilen uçtan-uca ürün tanıma modelinin (SSD + SURF&BRISK&ORB + NR-CAlg) CA değeriyle [50]'den daha yüksek bir doğruluk elde edilmiştir. Önerilen modelin PA ve PR değerleri [6] ve [36]'dan yüksek, PP değeri de [6]'dan oldukça yüksektir. Çalışma [6] ve önerilen yöntemle elde edilen performans değerleri karşılaştırıldığında, önerilen çok aşamalı yöntemin ürün tespiti ve sınıflandırması aşamalarının ortak olarak ele alındığı çalışma [6]'dan daha yüksek performans gösterdiği görülmektedir. Yine elde edilen sonuçlara göre, ürün paketleri üzerinde yer alan kelimeleri tespit ederek ürünleri tanımaya çalışmanın [50], diğer sonuçlara göre yeterince başarılı olmadığı görülmektedir. Önerilen yöntem SSD + SURF&BRISK&ORB + NR-CAlg, karşılaştırılan çalışmalar [6], [36] ve [50]'de elde edilen performans ölçümlerinin çoğunda galip gelmiştir.

Tablo 4.5 Önerilen yöntemin (SSD + SURF&BRISK&ORB + NR-CAlg) Grocery Products veri seti üzerindeki performansının yapılan çalışmalarla karşılaştırılması

	CA	PA	PP	PR
[6]	-	21,2	23,5	43,1
[36]	<u>84,6</u>	32,5	<u>57,0</u>	41,6
[50]	61,9	-	-	-
SSD + SURF&BRISK&ORB +NR-CAlg	76,4	<u>41,2</u>	39,4	<u>48,2</u>

4.4 Çalışma Zamanı Performansı Değerlendirmesi

Otomatik bir market raflarında ürün tanıma sisteminin gerçek zamanlı olarak çalışması önemli bir ihtiyaçtır. Çünkü üretici, tedarikçi ve müşteri açısından sorunlarına doğru şekilde cevap verilebilmesi kadar cevap verme süresi oldukça önemlidir. Bu tez çalışmasında ürün tanıma performansının yanı sıra GP-20 veri seti [15] üzerinde çalışma süresinin değerlendirilmesi de dikkate alınmıştır.

Ürün algılama aşamasında, üründen bağımsız algılama süreci sayesinde SSD'yi eğitirken eğitim süresi göz ardı edilebilmektedir. Çünkü SSD'nin, her yeni ürün eklendiğinde ya da ürünlerin ambalajı değiştiğinde tekrar eğitilmesi gerekmez. Ürün algılama aşamasının ortalama test süresi ise GP-20'deki 71 raf görüntüsü üzerinde 0,37 saniye olarak tespit edilmiştir. Raf görsellerinde yer alan ürünlerin sayısı ve ürünlerin boyutları ürün algılama süresini doğrudan etkilemektedir.

Ürün sınıflandırma aşamasında SURF, BRISK ve ORB'nin hibrit kullanımı önemli bir zaman kaybına yol açmamıştır. GP-20 veri seti 20 ürün şablon görüntüsünün her bir özellik çıkarma yöntemi ve önerilen hibrit yöntem ile elde edilen çalışma süreleri Tablo 4.10'da gösterilmiştir. Her yeni ürün eklendiğinde, bu ürün için en ayırt edici özelliğin SURF, BRISK ve ORB'den hangisi olduğunu tespit etmek için üç özellik çıkarma yöntemi de kullanılır. Ancak ürün için hangi özelliğin en belirgin olduğuna karar verildikten sonraki süreçte, yeni raf görselleri üzerinde test edilirken yalnızca bir özellik çıkarma yöntemiyle (o ürün için belirlenen en ayırt edici özellik

ile) devam edilir. Bu nedenle SURF, BRISK ve ORB'nin hibrit kullanımı, SURF'ün tek başına kullanılmasından daha hızlıdır. Ayrıca GP-20 veri setinin kullanıldığı başka bir çalışmada [15] bu süre 10 eğitim örneği için ortalama 1 saniyedir.

Tablo 4.6 Ürün sınıflandırma aşamasında kullanılan yöntemlerin çalışma sürelerinin GP-20 veri seti üzerinde karşılaştırılması

SURF	BRISK	ORB	SURF&BRISK&ORB
0,2094 s	0,0669 s	0,0433 s	0,0934 s

Ayrıca, çalışma zamanı performansını etkileyen en önemli faktör olan hesaplama karmaşıklığı için de SURF, BRISK ve ORB'nin ayrı ayrı ve hibrit kullanımlarının sonuçları Tablo 4.11'de verilmiştir. Her özellik çıkarma yönteminin hesaplama karmaşıklığını etkileyen farklı parametreler mevcuttur. SURF'ün hesaplama karmaşıklığı görüntüde tespit edilen ilgi noktalarının sayısına ve özellik tespiti için kullanılan Gauss çekirdeğinin boyutuna bağlıdır. BRISK'in hesaplama karmaşıklığı, ölçek uzay piramidinde kullanılan oktav ve katmanların sayısı, tespit edilen anahtar noktaların sayısı ve özellik tanımlayıcının boyutu dahil olmak üzere çeşitli faktörlerden etkilenir. ORB'nin hesaplama karmaşıklığı da tespit edilen anahtar noktaların sayısı, görüntünün boyutu ve tanımlayıcı eşleştirme sürecinin karmaşıklığı gibi faktörlere bağlıdır.

Tablo 4.7 Ürün sınıflandırma aşamasında kullanılan yöntemlerin hesaplama karmaşıklığı

SURF	BRISK	ORB	SURF&BRISK&ORB
$O(n \log n)$	$O(n)$	$O(n)$	$O(n \log n)$

Burada n değeri, görüntüde tespit edilen anahtar noktaların sayısını göstermek üzere genel olarak, SURF'ün zaman karmaşıklığı $O(n \log n)$, BRISK ve ORB'nin zaman karmaşıklığı $O(n)$ olarak temsil edilir. Bu durumda bu özelliklerin hibrit kullanımı için en kötü durum analizi de SURF ile aynı şekilde $O(n \log n)$ olacaktır. Bu nedenle, bu çalışmada önerdiğimiz SURF, BRISK ve ORB'nin hibrit kullanımı sadece SURF'ün kullanılması ile aynı hesaplama karmaşıklığına sahipken, sadece BRISK veya sadece ORB kullanımından daha çok hesaplama karmaşıklığına sahiptir.

5

SONUÇ

Market raflarındaki ürünlerin tanınması ile planogram uyumluluğu, stok takibi ve müşteri desteği gibi önemli problemlerin çözümüne katkı sağlanmaktadır. Başarılı bir otomatik ürün tanıma sistemi hem perakendecilere hem de müşterilere zamandan ve maliyetten tasarruf sağlamaktadır. Ürün çeşitliliği, sınıflar arası benzerliğin fazla olması, veri toplamının zorluğu, farklı karmaşıklığa sahip raf tasarımları gibi nedenler market raflarında ürün tanımayı zorlaştırmaktadır. Bu zorlukların üstesinden gelmek için çeşitli bilgisayarla görme ve makine öğrenmesi yöntemleri kullanılarak çözümler üretilmektedir.

Bu tez çalışmasında market raf görüntülerindeki ürünlerin çok aşamalı uçtan-uca tanınması için hibrit bir yöntem önerilmiştir. Ürün tanıma süreci, ürün konumlarının tespiti ile başlar; daha sonra süreç ürünleri sınıflandırır ve son olarak da iyileştirmeye sonuculanır.

Çok aşamalı uçtan-uca ürün tanıma sisteminin ilk aşaması olan ürün konumlarının tespitinde raflardaki her bir mevcut ürünün potansiyel bölgeleri belirlenir. Ürün tespit aşaması için dört farklı yöntem kullanılır: (1) Geleneksel bilgisayarla görme yaklaşımı (2) öğrenmeye dayalı bir ACF detektörü (3) DNN tabanlı YOLOv4 ve (4) DNN tabanlı SSD. Bu üç farklı temeldeki yaklaşım kullanılarak, ürün tanıma probleminin ilk aşaması olan nesne konumlarının tespiti için hangisinin daha uygun olduğu gösterilmiştir. Aşama-1 için SSD kullanımı, test edilen diğer ürün tespiti yöntemleri geleneksel bilgisayarla görme yaklaşımı, ACF ve YOLOv4'ten daha etkilidir. Ayrıca, SSD kullanımı üründen bağımsız bir ürün tespit süreci sağladığından sistemin yeni eklenen ya da ambalajı değişen ürünlerle tekrar eğitimi gerektirmez.

Sistemin ikinci aşaması olan ürün sınıflandırma aşamasında ise aşama-1'den elde edilen bölge önerileri ile ürün şablon görüntüleri özellik eşleştirmesi yaparak sınıflandırılır. Özellik çıkarımı için SURF, BRISK, ORB ve bu özelliklerin hibrit kullanımı ile gerçekleştirilmiştir. Aşama-2 için SURF, BRISK, ORB özelliklerinin hibrit kullanımının bu özelliklerin ayrı ayrı kullanılmasından daha iyi sonuçlar verdiği görülmüştür.

Son aşama olan iyileştirme aşamasında raf görüntüsünde aynı ürün için elde edilen çoklu bölge önerileri indirgenerek nihai sonuca ulaşılır. Bu aşamada kullanılan kümeleme algoritmasına komşuluk ilişkisinin eklenmesi ile iyileştirme algoritması geliştirilmiş ve performansı artırıcı etki göstermiştir.

Çalışma sonucunda market raflarındaki ürünlerin görüntüden tanınması için önerilen yöntem birinci aşama için SSD, ikinci aşama için SURF, BRISK, ORB'nin hibrit kullanımı ve komşuluk ilişkili kümeleme algoritmasıdır. Önerilen çok aşamalı yöntemin farklı ölçeklerde ve özelliklerdeki veri setleri üzerinde alınan deneysel sonuçları, farklı metriklerin çoğunda önceki çalışmalardan daha yüksek performans elde etmiştir. Ayrıca önerilen çok aşamalı yöntem, ürün tespiti ve sınıflandırması aşamalarının ortak olarak ele alındığı çalışmadan daha yüksek performans göstermiştir. Çalışma zamanı performans değerlendirmesine göre de gerçek zamanlı sistem gereksinimlerini karşılamıştır.

Market raflarındaki ürünlerin tanınması problemi hem akademik hem de endüstriyel açıdan geniş araştırma çalışmaları yapılan önemli bir problem olarak devam etmektedir. Derin öğrenme tabanlı yöntemlerin başarısını kanıtladığı birçok nesne tespiti ve sınıflandırılması alanı olmasına rağmen, market ürünleri veri setleri bu ağları eğitmek için yeterli değildir. Daha derin modellerle daha büyük veri setleri üzerinde çalışmak gereklidir. Bu nedenle, market ürünleri veri setlerini büyütme ve etiketli veri elde etmek için yeni yaklaşımların ve yöntemlerin geliştirilmesine ihtiyaç duyulmaktadır. Geleneksel yöntemlerin yanı sıra yapılacak çalışmalarda, grafik tabanlı modeller kullanılarak ürün tanıma sisteminin başarısı artırılabilir. Grafik tabanlı modeller, ürünler arasındaki bağlantıları ve ilişkileri temsil etmek için grafik yapısını kullanarak raf dizilimlerindeki ürünlerin konumları ve komşuluk ilişkilerini modelleyebilir. Böylece, bir ürünün sık sık yan yana veya birbirine yakın olduğu diğer ürünlerle ilişkili olabileceği veya belirli bir ürünün bulunduğu konumun, o ürünün türü veya kategorisi ile ilişkili olabileceği gibi bilgiler çıkarılabilir. Market raflarında ürün tanıma probleminin zorluklarının üstesinden gelmede kullanılan geleneksel yöntemlerin tek başına yeterli olmadığı durumlarda bu grafik tabanlı bilgiler ile daha hassas ve doğru tahminler yapılması sağlanabilir.

- [1] C. Ebster, *Store design and visual merchandising: Creating store space that encourages buying*. Business Expert Press, 2011.
- [2] M. Shapiro, “Executing the best planogram,” *Professional Candy Buyer*, Norwalk, CT, USA, 2009.
- [3] T. W. Gruen, D. S. Corsten, and S. Bharadwaj, *Retail out-of-stocks: A worldwide examination of extent, causes and consumer responses*. Grocery Manufacturers of America Washington, DC, 2002.
- [4] R. Berger, “Optimal Shelf Availability: Increasing Shopper Satisfaction at the Moment of Truth,” <http://ecr-community.org/wp-content/uploads/2016/10/ecr-europe-osa-optimal-shelf-availability.pdf>, Oct. 2016.
- [5] E. Goldman, R. Herzig, A. Eisenschtat, J. Goldberger, and T. Hassner, “Precise detection in densely packed scenes,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2019-June, pp. 5222–5231, 2019, doi: 10.1109/CVPR.2019.00537.
- [6] M. George and C. Floerkemeier, “LNCS 8690 - Recognizing Products: A Per-exemplar Multi-label Image Classification Approach,” 2014.
- [7] P. Dollar, R. Appel, S. Belongie, and P. Perona, “Fast feature pyramids for object detection,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 8, pp. 1532–1545, 2014, doi: 10.1109/TPAMI.2014.2300479.
- [8] C. Wang and H. M. Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection”.
- [9] W. Liu *et al.*, “SSD: Single shot multibox detector,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9905 LNCS, pp. 21–37, 2016, doi: 10.1007/978-3-319-46448-0_2.
- [10] H. Bay, T. Tuytelaars, and L. Van Gool, “LNCS 3951 - SURF: Speeded Up Robust Features,” 2006.
- [11] S. Leutenegger, M. Chli, and R. Y. Siegwart, “BRISK: Binary Robust invariant scalable keypoints,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2011, pp. 2548–2555. doi: 10.1109/ICCV.2011.6126542.
- [12] E. Rosten and T. Drummond, “Machine learning for high-speed corner detection,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 3951 LNCS, no. June, pp. 430–443, 2006, doi: 10.1007/11744023_34.
- [13] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, “BRIEF: Binary Robust Independent Elementary Features ★”.
- [14] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, “ORB: An efficient alternative to SIFT or SURF,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2011, pp. 2564–2571. doi:

10.1109/ICCV.2011.6126544.

- [15] A. Franco, D. Maltoni, and S. Papi, "Grocery product detection and recognition," *Expert Syst. Appl.*, vol. 81, pp. 163–176, Sep. 2017, doi: 10.1016/j.eswa.2017.02.050.
- [16] A. Tonioni and L. Di Stefano, "Product recognition in store shelves as a sub-graph isomorphism problem," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10484 LNCS, pp. 682–693, 2017, doi: 10.1007/978-3-319-68560-1_61.
- [17] G. Varol and R. S. Kuzu, "Toward retail product recognition on grocery shelves," no. March, 2015, doi: 10.1117/12.2179127.
- [18] W. P. Fernandez, Y. Xian, and Y. Tian, "Image-Based Barcode Detection and Recognition to Assist Visually Impaired Persons," *2017 IEEE 7th Annu. Int. Conf. CYBER Technol. Autom. Control. Intell. Syst. CYBER 2017*, pp. 1241–1245, 2018, doi: 10.1109/CYBER.2017.8446388.
- [19] V. Kulyukin and A. Kutianawala, "From ShopTalk to ShopMobile: vision-based barcode scanning with mobile phones for independent blind grocery shopping," *Proc. 2010 Rehabil. Eng. Assist. Technol. Soc. North Am. Conf. (RESNA 2010), Las Vegas, NV*, no. 703, pp. 1–5, 2010, [Online]. Available: http://digital.cs.usu.edu/~vkulyukin/vkweb/pubs/RESNA2010_VKulyukin1.pdf
- [20] C. Condea, F. Thiesse, and E. Fleisch, "RFID-enabled shelf replenishment with backroom monitoring in retail stores," *Decis. Support Syst.*, vol. 52, no. 4, pp. 839–849, 2012, doi: 10.1016/j.dss.2011.11.018.
- [21] C. Metzger, F. Thiesse, S. Gershwin, and E. Fleisch, "The impact of false-negative reads on the performance of RFID-based shelf inventory control policies," *Comput. Oper. Res.*, vol. 40, no. 7, pp. 1864–1873, 2013, doi: 10.1016/j.cor.2013.02.001.
- [22] M. Wolbitsch, T. Hasler, M. Goller, C. Gutl, S. Walk, and D. Helic, "RFID in the Wild - Analyzing Stocktake Data to Determine Detection Probabilities of Products," *2019 6th Int. Conf. Internet Things Syst. Manag. Secur. IOTSMS 2019*, pp. 251–258, 2019, doi: 10.1109/IOTSMS48152.2019.8939247.
- [23] M. F. M. Busu, I. Ismail, M. F. Saaied, and S. M. Norzeli, "Auto-checkout system for retails using Radio Frequency Identification (RFID) technology," *Proc. - 2011 IEEE Control Syst. Grad. Res. Colloquium, ICSGRC 2011*, pp. 193–196, 2011, doi: 10.1109/ICSGRC.2011.5991855.
- [24] P. L. Grieco, M. W. Gozzo, and C. J. Long, *Behind Bars: Bar Coding Principles and Applications*. PT Publications, 1989. [Online]. Available: <https://books.google.com.tr/books?id=yD8WAAAACAAJ>
- [25] M. Murad, A. Rehman, A. A. Shah, S. Ullah, M. Fahad, and K. M. Yahya, "RFAIDE - An RFID based navigation and object recognition assistant for visually impaired people," *2011 7th Int. Conf. Emerg. Technol. ICET 2011*, no. September, 2011, doi: 10.1109/ICET.2011.6048486.
- [26] L. McCathie, "The advantages and disadvantages of barcodes and radio frequency identification in supply chain management," *Univ. Wollongong*

Thesis Collect., p. 125, 2004.

- [27] F. Maulana, Nixon, R. P. Putra, and N. Hanafiah, "Self-Checkout System Using RFID (Radio Frequency Identification) Technology: A Survey," no. October, pp. 273–277, 2021, doi: 10.1109/iccsai53272.2021.9609762.
- [28] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, 2017, doi: 10.1109/TPAMI.2016.2577031.
- [29] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 6517–6525, 2017, doi: 10.1109/CVPR.2017.690.
- [30] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "[ConvNeXt_CVPR22] A ConvNet for the 2020s," *Cvpr*, pp. 11976–11986, 2022, [Online]. Available: <http://arxiv.org/abs/2201.03545>
- [31] M. Merler, C. Galleguillos, and S. Belongie, "Recognizing groceries in situ using in vitro training data," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2007, doi: 10.1109/CVPR.2007.383486.
- [32] T. Winlock, E. Christiansen, and S. Belongie, "Toward real-time grocery detection for the visually impaired," *2010 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. - Work. CVPRW 2010*, pp. 49–56, 2010, doi: 10.1109/CVPRW.2010.5543576.
- [33] L. Karlinsky, J. Shtok, Y. Tzur, and A. Tzadok, "Fine-grained recognition of thousands of object categories with single-example training," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 965–974, 2017, doi: 10.1109/CVPR.2017.109.
- [34] I. Baz, E. Yoruk, and M. Cetin, "Context-Aware Confidence Sets for Fine-Grained Product Recognition," *IEEE Access*, vol. 7, pp. 76376–76393, 2019, doi: 10.1109/ACCESS.2019.2921994.
- [35] F. De Feyter, "Joint Training of Product Detection and Recognition Using Task-Specific Datasets," vol. 5, no. Visigrapp, pp. 715–722, 2023, doi: 10.5220/0011725100003417.
- [36] E. Yörük, K. T. Öner, and C. B. Akgül, "An efficient Hough transform for multi-instance object recognition and pose estimation," *Proc. - Int. Conf. Pattern Recognit.*, vol. 0, pp. 1352–1357, 2016, doi: 10.1109/ICPR.2016.7899825.
- [37] G. Varol, "Product Placement Detection Based on Image Processing," no. April, 2014, doi: 10.1109/SIU.2014.6830408.
- [38] X. Pan *et al.*, "Dynamic refinement network for oriented and densely packed object detection," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 11204–11213, 2020, doi: 10.1109/CVPR42600.2020.01122.
- [39] Ü. Gökdag, "Planogram Matching Control in Grocery Products by Image Processing," no. May, 2016, doi: 10.1109/SIU.2016.7495972.
- [40] M. M. Srivastava, "Bag of tricks for retail product image classification," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect.*

- Notes Bioinformatics*), vol. 12131 LNCS, pp. 71–82, 2020, doi: 10.1007/978-3-030-50347-5_8.
- [41] S. Kant, “Learning Gaussian Maps for Dense Object Detection,” pp. 1–13, 2020, [Online]. Available: <http://arxiv.org/abs/2004.11855>
 - [42] C. Wang, C. Huang, X. Zhu, and L. Zhao, “One-Shot Retail Product Identification Based on Improved Siamese Neural Networks,” *Circuits, Syst. Signal Process.*, vol. 41, no. 11, pp. 6098–6112, 2022, doi: 10.1007/s00034-022-02062-y.
 - [43] C. Xu, Y. Zheng, Y. Zhang, G. Li, and Y. Wang, “A method for detecting objects in dense scenes,” pp. 75–82, 2022.
 - [44] T. Tiwary and R. P. Mahapatra, “Enhancement in web accessibility for visually impaired people using hybrid deep belief network – bald eagle search,” 2023.
 - [45] Y. Zhang, L. Wang, R. Hartley, and H. Li, “Where ’ s the Weet-Bix ?,” pp. 800–801, 2007.
 - [46] P. Jund, N. Abdo, A. Eitel, and W. Burgard, “The Freiburg Groceries Dataset,” 2016, [Online]. Available: <http://arxiv.org/abs/1611.05799>
 - [47] A. Tonioni, E. Serra, and L. Di Stefano, “A deep learning pipeline for product recognition on store shelves,” *IEEE 3rd Int. Conf. Image Process. Appl. Syst. IPAS 2018*, pp. 25–31, 2018, doi: 10.1109/IPAS.2018.8708890.
 - [48] P. Selvam and J. A. S. Koilraj, “A Deep Learning Framework for Grocery Product Detection and Recognition,” *Food Anal. Methods*, vol. 15, no. 12, pp. 3498–3522, 2022, doi: 10.1007/s12161-022-02384-2.
 - [49] F. Chen *et al.*, “Unitail: Detecting, Reading, and Matching in Retail Scene,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 13667 LNCS, pp. 705–722, 2022, doi: 10.1007/978-3-031-20071-7_41.
 - [50] M. George, D. Mircic, G. Sörös, C. Floerkemeier, and F. Mattern, “Fine-Grained Product Class Recognition for Assisted Shopping,” *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2015-Febru, pp. 546–554, 2015, doi: 10.1109/ICCVW.2015.77.
 - [51] E. Frontoni, M. Contigiani, G. Ribighini, and I. Dii, “A heuristic approach to evaluate occurrences of products for the planogram maintenance,” *2014 IEEE/ASME 10th Int. Conf. Mechatron. Embed. Syst. Appl.*, no. Dii, pp. 1–6, doi: 10.1109/MESA.2014.6935615.
 - [52] S. Liu and H. Tian, “Planogram Compliance Checking using Recurring Patterns,” *2015 IEEE Int. Symp. Multimed.*, pp. 27–32, 2015, doi: 10.1109/ISM.2015.72.
 - [53] R. O. Duda and P. E. Hart, “Use of the Hough Transformation To Detect Lines and Curves in Pictures,” vol. 15, no. 1, 1972, doi: 10.1145/361237.361242.
 - [54] N. Dalal and B. Triggs, “Histograms of Oriented Gradients for Human Detection”.
 - [55] P. Viola and M. Jones, “Rapid Object Detection using a Boosted Cascade of

- Simple Features,” pp. 1–8, 2001.
- [56] D. G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints,” vol. 60, no. 2, pp. 91–110, 2004.
 - [57] G. H. Å, Q. Zhu, and C. Siew, “Extreme learning machine : Theory and applications,” vol. 70, pp. 489–501, 2006, doi: 10.1016/j.neucom.2005.12.126.
 - [58] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, “Focal Loss for Dense Object Detection,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, 2020, doi: 10.1109/TPAMI.2018.2858826.
 - [59] J. Redmon and A. Farhadi, “YOLOv3: An Incremental Improvement,” 2018, [Online]. Available: <http://arxiv.org/abs/1804.02767>
 - [60] X. Zhou, D. Wang, and P. Krähenbühl, “Objects as Points,” 2019, [Online]. Available: <http://arxiv.org/abs/1904.07850>
 - [61] Z. Cai, “Cascade R-CNN : Delving into High Quality Object Detection”.
 - [62] G. Koch, “Siamese Neural Networks for One-shot Image Recognition,” 2011.
 - [63] K. Mikolajczyk and C. Schmid, “Scale & Affine Invariant Interest Point Detectors,” vol. 60, no. 1, pp. 63–86, 2004.
 - [64] A. Krizhevsky and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” pp. 1–9.
 - [65] Y. Jia *et al.*, “Caffe : Convolutional Architecture for Fast Feature Embedding *”.
 - [66] R. Smith, “An Overview of the Tesseract OCR Engine,” 2005.
 - [67] D. J. C. MacKay, “Information-Based Objective Functions for Active Data Selection,” *Neural Comput.*, vol. 4, no. 4, pp. 590–604, 1992, doi: 10.1162/neco.1992.4.4.590.
 - [68] L. Barrington, T. K. Marks, J. H. W. Hsiao, and G. W. Cottrell, “Nimble: A kernel density model of saccade-based visual memory,” *J. Vis.*, vol. 8, no. 14, pp. 1–14, 2008, doi: 10.1167/8.14.17.
 - [69] B. D. Lucas, “An Iterative Image Registration Technique with an Application to Stereo Vision,” no. April 1981, 2016.
 - [70] B. Yao, A. Khosla, and L. Fei-Fei, “Combining randomization and discrimination for fine-grained image categorization,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 1577–1584, 2011, doi: 10.1109/CVPR.2011.5995368.
 - [71] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, “Return of the devil in the details: Delving deep into convolutional nets,” *BMVC 2014 - Proc. Br. Mach. Vis. Conf. 2014*, pp. 1–11, 2014, doi: 10.5244/c.28.6.
 - [72] Y. Wei, S. Tran, S. Xu, B. Kang, and M. Springer, “Deep Learning for Retail Product Recognition: Challenges and Techniques,” *Comput. Intell. Neurosci.*, vol. 2020, 2020, doi: 10.1155/2020/8875910.
 - [73] “Detecting retail products in situ using CNN without human effort labeling

Wei Yi * , Yaoran Sun, Tao Ding, Sailing He Center for Optical and Electromagnetic Research, Zhejiang University 310058, P.R. China”.

- [74] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–14, 2015.
- [75] G. Jocher *et al.*, “ultralytics/yolov5: Initial Release.” Zenodo, 2020. doi: 10.5281/zenodo.3908560.
- [76] K. He, “Deep Residual Learning for Image Recognition”.
- [77] R. Litman, O. Anschel, S. Tsiper, R. Litman, S. Mazor, and R. Manmatha, “Scatter: Selective context attentional scene text recognizer,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 11959–11969, 2020, doi: 10.1109/CVPR42600.2020.01198.
- [78] L. Huang, Y. Yang, Y. Deng, and Y. Yu, “DenseBox: Unifying Landmark Localization with End to End Object Detection,” pp. 1–13, 2015, [Online]. Available: <http://arxiv.org/abs/1509.04874>
- [79] J. Robinson and S. Assignment, “The hungarian method for the assignment problem,” *Nav. Res. Logist. Q.*, vol. 2, pp. 14–16, 1950, [Online]. Available: <https://onlinelibrary.wiley.com/doi/epdf/10.1002/nav.3800020109>
- [80] W. Geng *et al.*, “Fine-Grained Grocery Product Recognition by One-Shot Learning,” vol. 2, pp. 1706–1714, 2018.
- [81] Y. Cai, L. Wen, L. Zhang, D. Du, and W. Wang, “Rethinking Object Detection in Retail Stores,” *35th AAAI Conf. Artif. Intell. AAAI 2021*, vol. 2A, pp. 947–954, 2021, doi: 10.1609/aaai.v35i2.16178.
- [82] A. Osokin, D. Sumin, and V. Lomakin, “OS2D: One-Stage One-Shot Object Detection by Matching Anchor Features,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12360 LNCS, pp. 635–652, 2020, doi: 10.1007/978-3-030-58555-6_38.
- [83] R. Girshick, J. Donahue, T. Darrell, J. Malik, and U. C. Berkeley, “Rich feature hierarchies for accurate object detection and semantic segmentation,” 2012.
- [84] C. Du and D. Sun, “Comparison of three methods for classification of pizza topping using different colour space transformations,” vol. 68, pp. 277–287, 2005, doi: 10.1016/j.jfoodeng.2004.05.044.
- [85] P. F. Alcantarilla and A. Bartoli, “Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces”.
- [86] S. Xie, R. Girshick, and P. Doll, “Aggregated Residual Transformations for Deep Neural Networks”.
- [87] D. G. Stork, “Pattern Classification,” no. April, 2016.
- [88] L. Barrington, T. K. Marks, J. H. Hsiao, and G. W. Cottrell, “NIMBLE: a kernel density model of saccade-based visual memory,” *J. Vis.*, vol. 8, no. 14, pp. 17.1–14, Nov. 2008, doi: 10.1167/8.14.17.
- [89] L. E. O. Breiman, “Random Forests,” pp. 5–32, 2001.
- [90] A. Neubeck, “Efficient Non-Maximum Suppression,” 2006.

- [91] C. Wang, A. Bochkovskiy, and H. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," pp. 1–15.
- [92] G. Melek, "Object Detection in Shelf Images with YOLO," pp. 1–5, 2019.
- [93] B. J. Friedman, T. Hastie, and R. Tibshirani, "ADDITIVE LOGISTIC REGRESSION: A STATISTICAL VIEW OF BOOSTING," *Ann. Stat.*, vol. 28, no. 2, pp. 337–407, 2000.
- [94] B. T. Bastian and C. V. Jiji, "Integrated feature set using aggregate channel features and histogram of sparse codes for human detection," pp. 2931–2944, 2020.
- [95] D. Yi, J. Su, and W. Chen, "Locust Recognition and Detection via Aggregate Channel Features".
- [96] A. Zhao *et al.*, "An Effective Method Based on ACF for Aircraft Detection in Remote Sensing Images," vol. 14, no. 5, pp. 744–748, 2017.
- [97] F. A. Hermawati, "Combination of Aggregated Channel Features (ACF) Detector and Faster R-CNN to Improve Object Detection Performance in Fetal Ultrasound Images," vol. 11, no. 6, 2018, doi: 10.22266/ijies2018.1231.07.
- [98] R. Girshick, "Fast R-CNN".
- [99] W. Y. Hsu and W. Y. Lin, "Adaptive Fusion of Multi-Scale YOLO for Pedestrian Detection," *IEEE Access*, vol. 9, pp. 110063–110073, 2021, doi: 10.1109/ACCESS.2021.3102600.
- [100] S. Basovnik, L. Mach, A. Mikulik, and D. Obdrzalek, "Detecting Scene Elements Using Maximally Stable Colour Regions".
- [101] Z. Saleem, "A Comparative Analysis of SIFT , SURF , KAZE , AKAZE , ORB , and BRISK," *2018 Int. Conf. Comput. Math. Eng. Technol.*, pp. 1–10, 2018, doi: 10.1109/ICOMET.2018.8346440.
- [102] E. Mair, G. D. Hager, D. Burschka, M. Suppa, and G. Hirzinger, "Computer Vision -- ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part II," *Adapt. Generic Corner Detect. Based Accel. Segm. Test*, pp. 183–196, 2010.
- [103] Jose Gabriel Islas Montero, "Multiclass Confusion Matrix for Object Detection." <https://www.tenyks.ai/blog/multiclass-confusion-matrix-for-object-detection>, 12.05.2023.

TEZDEN ÜRETİLMİŞ YAYINLAR

Konferans Bildirileri

1. C. G. Melek, E. B. Sonmez and S. Albayrak, "A survey of product recognition in shelf images," *2017 International Conference on Computer Science and Engineering (UBMK)*, Antalya, Turkey, 2017, pp. 145-150, doi: 10.1109/UBMK.2017.8093584.
2. C. G. Melek, E. B. Sonmez and S. Albayrak, "Object Detection in Shelf Images with YOLO," *IEEE EUROCON 2019 -18th International Conference on Smart Technologies*, Novi Sad, Serbia, 2019, pp. 1-5, doi: 10.1109/EUROCON.2019.8861817.

Makaleler

1. C. G. Melek, E. Battini Sonmez, H. Ayrar and S. Varli, "Development of a Hybrid Method for Multi-Stage End-to-End Recognition of Grocery Products in Shelf Images," *Electronics* 2023, 12, 3640. <https://doi.org/10.3390/electronics12173640>.