

T.C.
GEBZE TEKNİK ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

OTONOM SÜRÜŞ İÇİN ANLAMSAL BÖLÜTLEMEDE
ALAN UYARLAMA

YAKUP YILDIZ

YÜKSEK LİSANS TEZİ
ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI

DANIŞMAN: PROF. DR. KORAY KAYABOL

TEMMUZ 2025

T.C.
GEBZE TEKNİK ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

OTONOM SÜRÜŞ İÇİN ANLAMSAL BÖLÜTLEMEDE
ALAN UYARLAMA

YAKUP YILDIZ

YÜKSEK LİSANS TEZİ
ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI

DANIŞMAN: PROF. DR. KORAY KAYABOL

TEMMUZ 2025

T.R.
GEBZE TECHNICAL UNIVERSITY
GRADUATE SCHOOL

DOMAIN ADAPTATION
IN SEMANTIC SEGMENTATION
FOR AUTONOMOUS DRIVING

YAKUP YILDIZ

A THESIS OF MASTER OF SCIENCE
DEPARTMENT OF ELECTRONICS ENGINEERING

ADVISOR: PROF. DR. KORAY KAYABOL

JULY 2025

YÜKSEK LİSANS JÜRİ ONAY FORMU

GTÜ Lisansüstü Eğitim Enstitüsü Yönetim Kurulunun 30/06/2025 tarih ve 2025/34 sayılı kararıyla oluşturulan jüri tarafından 09/07/2025 tarihinde tez savunma sınavı yapılan Yakup YILDIZ'ın tez çalışması Elektronik Mühendisliği Anabilim Dalında YÜKSEK LİSANS tezi olarak kabul edilmiştir.

JÜRİ

ÜYE

(TEZ DANIŞMANI) : Prof. Dr. Koray KAYABOL

ÜYE

: Dr. Öğr. Üyesi Köksal HOCAOĞLU

ÜYE

: Prof. Dr. Erchan APTOULA

ONAY

Gebze Teknik Üniversitesi Lisansüstü Eğitim Enstitüsü Yönetim Kurulunun

...../...../..... tarih ve/..... sayılı kararı.

İMZA/MÜHÜR



*Aileme
ve
biricik eşime*

ÖZET

Anlamsal bölütleme, bir görüntüdeki her bir pikselin ait olduğu kategoriye göre sınıflandırılması işlemidir. Bu yöntem; robotik, otonom araçlar, medikal görüntüleme ve havacılık sistemleri gibi birçok alanda yaygın olarak kullanılmaktadır. Anlamsal bölütleme modellerinin eğitilebilmesi için yüksek çözünürlüklü ve etiketli verilere ihtiyaç duyulmaktadır. Ancak gerçek dünya verilerinin toplanması hem maliyetli hem de zaman alıcı bir süreçtir. Bu sebeple, bu tür verilerin toplanması için simülasyonlardan faydalanmak, teknolojinin sunduğu olanaklar doğrultusunda tercih edilebilir hale gelmiştir. Ancak, simülasyon verileri ile gerçek veriler arasında bulunan dağılımsal farkları azaltmak için alan uyarlama (domain adaptation) yöntemlerine başvurulmaktadır.

Bu tez çalışmasında, GTA5 video oyunu veri kümesi [1] kaynak alan olarak, Cityscapes veri kümesi [2] ise hedef alan olarak kullanılmıştır. Frekans tabanlı alan uyarlama yöntemlerinin anlamsal bölütleme üzerindeki etkileri incelenmiş ve karşılaştırmalı analizler yapılmıştır. İlk olarak, Fourier Alan Uyarlaması (Fourier Domain Adaptation - FDA) yöntemi [3] ele alınmış, bu yöntemden ilham alınarak Gauss Fourier Alan Uyarlaması (Gaussian Fourier Domain Adaptation - GFDA) [4] ve Frekans Temelli Optimal Stil Karıştırma (Frequency-based Optimal Style Mix - FOSMix) [5] yöntemleri geliştirilmiştir.

FDA ve GFDA yöntemlerinde sırasıyla sabit kare maske ve yumuşak geçişli öğrenilebilir Gauss maske kullanılarak %42.31 ve %42.03 mean intersection over union (mIoU) değeri elde edilmiştir.

Üçüncü yöntem olarak, aslında bir alan genelleştirme (domain generalization) yaklaşımı olan FOSMix yöntemi incelenmiştir. Bu yöntem kapsamında Hızlı Fourier Dönüşümü (FFT) kullanılarak üç farklı maskeleme stratejisi —Gauss maske, optimal U-Net[28] tabanlı maske ve sabit maske— uygulanmış ve sırasıyla %41.53, %34.10 ve %41.81 mIoU değerleri elde edilmiştir. Bunun yanı sıra, aynı yöntem Ayrık Kosinüs Dönüşümü (DCT) ile de değerlendirilmiş; klasik maske stratejisinde %31.10, U-Net tabanlı optimal maske stratejisinde ise %28.52 mIoU başarımı rapor edilmiştir.

Anahtar Kelimeler: Alan Uyarlama, Anlamsal Bölütleme, Fourier Alan Uyarlaması, Gauss Maske, FOSMix

ABSTRACT

Semantic segmentation refers to the process of classifying each pixel in an image according to its corresponding category. This technique is widely used in various fields such as robotics, autonomous vehicles, medical imaging, and aerospace systems. Training models for semantic segmentation requires high-resolution, well-annotated datasets. However, collecting real-world data is both costly and time-consuming. As an alternative, thanks to advancements in technology, it has become preferable to use simulations to generate such data. Nevertheless, since there is a distribution gap between synthetic and real-world data, domain adaptation techniques are employed to reduce this discrepancy.

In this thesis study, the GTA5 video game dataset [1] is used as the source domain, while the Cityscapes dataset [2] represents the target domain. The performance of frequency-based domain adaptation methods in semantic segmentation is analyzed and compared. Initially, the Fourier Domain Adaptation (FDA) method [3] is investigated, which inspired the development of Gaussian Fourier Domain Adaptation (GFDA) [4] and Frequency-based Optimal Style Mix (FOSMix) [5].

In the FDA and GFDA methods, a fixed square mask and a learnable smooth Gaussian mask are applied, respectively, resulting in mean intersection over union (mIoU) scores of 42.31% and 42.03%.

As a third approach, the FOSMix method, which is fundamentally a domain generalization technique, is explored. Within the scope of this method, training is conducted using the Fast Fourier Transform (FFT) with three types of masks: Gaussian mask, U-Net-based optimal mask [8], and fixed mask, achieving mIoU performances of 41.53%, 34.10%, and 41.81%, respectively. Furthermore, the same method is also evaluated using the Discrete Cosine Transform (DCT), where the classical mask strategy yields 31.10% and the U-Net-based optimal mask strategy achieves 28.52% mIoU.

Keywords: Domain Adaptation, Semantic Segmentation, Fourier Domain Adaptation, Gaussian Mask, FOSMix

TEŞEKKÜR

Bu tez çalışmasının yürütülmesinde, bilgi ve deneyimleriyle bana her zaman yol gösteren; akademik desteği, sabrı ve teşvik edici yaklaşımıyla araştırma sürecine büyük katkı sağlayan danışmanım Prof. Dr. Koray Kayabol'a içten teşekkür ederim.

Bu süreçte her zaman yanımda olan, anlayışı ve desteğiyle bana güç veren sevgili eşim Feyza'ya teşekkür ederim. Ayrıca, bana her koşulda destek olan, sevgileri ve güvenleriyle daima yanımda hissettiğim aileme de gönülden teşekkür ederim.

Eğitim hayatım boyunca yanımda olan, birlikte emek verdiğimiz, bu tezde adı geçmeyen tüm arkadaşlarıma da verdikleri destek, moral, motivasyon ve dayanışma için teşekkür ederim.

Akademik hayatımın şekillenmesinde rol model olan, sadece bilgileriyle değil, duruşları ve verdikleri ilhamla da yolumu aydınlatan tüm saygıdeğer hocalarıma en içten teşekkürlerimi sunarım.

İÇİNDEKİLER

	Sayfa
ÖZET	vi
ABSTRACT	vii
TEŞEKKÜR	viii
İÇİNDEKİLER	ix
SİMGELER ve KISALTMALAR DİZİNİ	xi
ŞEKİLLER DİZİNİ	xii
TABLolar DİZİNİ	xiii
1. GİRİŞ	1
2. LİTERATÜR TARAMASI	7
2.1. Derin Öğrenme	7
2.2. Anlamsal Bölütleme	11
2.2.1. Anlamsal Bölütleme Yaklaşımları	12
2.2.1.1. Bölge Temelli Anlamsal Bölütleme	12
2.2.1.2. Tümden Evrişimsel Ağ Temelli Anlamsal Bölütleme	13
2.2.1.3. Az Denetlenen Anlamsal Bölütleme	14
2.2.2. Anlamsal Bölütlemelerde Kullanılan Modeller ve Açıklamaları	14
2.2.2.1. SegNet Mimarisi	14
2.2.2.2. U-Net Mimarisi	15
2.2.2.3. DeepLab Mimarisi	16
2.2.3. Anlamsal Bölütlemelerde Kullanılan Metrikler	18
2.2.3.1. Mean Intersection Over Union	18
2.2.3.2. Piksel Doğruluğu (Pixel Accuracy)	18
2.2.3.3. Keskinlik (Precision)	19
2.2.3.4. Duyarlılık (Recall)	19
2.2.3.5. F1 Skoru	19
2.2.3.6. Dice Skoru	29
2.3. Alan Uyarlama	19
2.3.1. Alan Uyarlama Yöntemleri	20
2.3.1.1. Denetimli Alan Uyarlama	20
2.3.1.2. Yarı Denetimli Alan Uyarlama	20
2.3.1.3. Denetimsiz Alan Uyarlama	21
2.3.2. Özellik Seviyesi Alan Uyarlama	21
2.3.3. Çekişmeli Alan Uyarlama	21
2.3.4. Piksel Düzeyinde Uyarlama	22
2.3.5. Frekans Tabanlı Alan Uyarlama	22
2.3.5.1. Fourier Alan Uyarlama	22
2.3.5.2. Gauss Fourier Alan Uyarlama	24
2.3.5.3. Frekans Tabanlı Optimal Stil Karışımı	25
2.4. Veri Kümeleri	27
3. YÖNTEM	28
3.1. Veri Ön İşlemesi	28
3.2. Alan Uyarlama	29
3.2.1. Fourier Alan Uyarlaması	29
3.2.2. Gauss Fourier Alan Uyarlaması	30

3.2.3. Frekans Tabanlı Optimal Stil Karışımı	31
3.3. Model Eğitimi	32
3.3.1. Model Parametreleri	33
3.3.2. Öğrenme Oranı	33
3.3.3. Eniyileyiciler	34
3.3.4. Eğitim Adımları ve Batch Büyüklüğü	34
3.3.5. Model Kayıt Süreci	34
3.3.6. Kayıp Fonksiyonları	35
3.4. Model Değerlendirmesi ve Başarım Metrikleri	36
3.4.1. Test İşlemi	37
3.4.2. Başarım Metrikleri	37
4. TEST VE ANALİZ	38
4.1. Veri Kümeleri	38
4.1.1. Cityscapes	38
4.1.2. GTA5	39
4.2. Performans Kriterleri	40
4.3. Test Sonuçları	40
5. SONUÇLAR VE YORUMLAR	47
6. GELECEK ÇALIŞMALAR	49
KAYNAKLAR	50
ÖZGEÇMİŞ	54
TEZ METNİNDEN TÜRETİLEN YAYINLAR/SUNUMLAR	55

SİMGELER ve KISALTMALAR DİZİNİ

CNN	: Convolutional Neural Network
DA	: Domain Adaptation
FCN	: Fully Connected Network
FM	: Full Mix
FDA	: Fourier Domain Adaptation
FOSMix	: Frequency-based Style Mixing
GFDA	: Gaussian Fourier Domain Adaptation
OM	: Optimal Mix
R-CNN	: Region Based Convolutional Neural Networks
ReLU	: Rectified Linear Unit



ŞEKİLLER DİZİNİ

	<u>Sayfa</u>
Şekil 2.1: Örnek evrişimsel sinir ağı [13].	8
Şekil 2.2: Maksimum havuzlama ve ortalama havuzlama [12].	9
Şekil 2.3: Basamak fonksiyonu [14].	10
Şekil 2.4: Doğrusal fonksiyon [14].	10
Şekil 2.5: Sigmoid fonksiyonu [14].	11
Şekil 2.6: ReLU Fonksiyonu [14].	11
Şekil 2.7: SegNet mimarisi [27].	15
Şekil 2.8: U-net mimarisi [28].	16
Şekil 2.9: DeepLab örnek akış şeması.	17
Şekil 2.10: FOSMix alan uyarlama akışı şeması [5].	26
Şekil 3.1: Veri ön işleme Akış diyagramı.	29
Şekil 3.2: Kare maske.	30
Şekil 3.3: Gauss maskesi.	31
Şekil 3.4: FOSMix optimal maske.	32
Şekil 3.5: Model akış şeması.	33
Şekil 4.1: Cityscapes örnek görüntü.	38
Şekil 4.2: Cityscapes örnek ground-truth.	38
Şekil 4.3: GTA5 örnek görüntüler	39
Şekil 4.4: GTA5 örnek ground-truth	39

TABLÖLER DİZİNİ

	<u>Sayfa</u>
Tablo 4.1: Yüzdesel sınıf dağılımları.	39
Tablo 4.2: IoU metrik sonuçları.	41
Tablo 4.3: Model tahmin görüntüleri ve ground-truth.	43
Tablo 4.4: Model tahmin görüntüleri.	43
Tablo 4.5: Geliştirilen yöntemlere ait metrik sonuçları.	44
Tablo 4.6: Yürütme süresi karşılaştırması.	45
Tablo 4.7: Bellek tüketimi karşılaştırması.	46

1. GİRİŞ

Bilgisayar bilimlerinin yayılması ile, görüntü işleme, önemli bir araştırma alanı olarak ilgileri üzerine çekmiştir. Görüntü işlemenin ilk dönemlerinde kullanılan yöntemlerden olan filtreleme, eşikleme ve kenar tespiti gibi yöntemler, genellikle görüntünün kalitesini iyileştirme veya nesne tanıma görevleri için kullanılmıştır. İlerleyen dönemlerde teknolojinin hızla gelişimi ve teknoloji alanında kullanılan cihazların işlem kapasitesinin artması, yapay zeka ve makine öğrenmesi alanlarında yüksek işlem hacimli çalışmaların yapılmasına olanak sağlamıştır. Bu çalışmalar neticesinde elde edilen bilgiler ile görüntü işleme dünyasında var olan problemlerin çözümü için yeni yollar aralanmış oldu.

ImageNet yarışması ile geliştirilen AlexNet evrimsel sinir ağının gösterdiği başarı derin öğrenme alanında bir mihenk taşı olarak bilinmektedir[6]. ImageNet yarışmasının sonucunda yapay zeka alanında derin öğrenme modellerinin yüksek hacimli ve kompleks veri kümeleri ile olağanüstü performans sergilediği gözükmiştir. Bu tarz ardı sıra gelen başarılı gelişmeler, görüntü işlemede önemli ve karmaşık problemler olan görüntü sınıflandırma, obje tanıma, anlamsal bölütleme, yüz tanıma ve parmak izi tanıma gibi yüksek veri hacmi ile eğitilmesi gereken alanlarda daha önce elde edilmemiş başarılarla öncülük etmiştir.

Derin öğrenmede elde edilen bu başarılı sonuçlar sebebiyle anlamsal bölütleme alanında yapılan araştırmalar hız kazanmıştır. Bunun sebebi derin öğrenme ile görüntülerden gözükmeyen ama var olan birçok öznelik veya başka bir deyişle özellik çıkartımı mümkün hale gelmiştir. Anlamsal bölütleme problemi, görüntüdeki her pikselin hangi anlamsal sınıfa ait olduğunun belirlenmesini içeren detaylı bir sınıflandırma sürecidir. Anlamsal bölütleme sayesinde görüntüde bulunan objelerin sınırları hassas bir belirlenebilmektedir. Bu sebeple otonom sürüş destekli sistemler, medikal görüntüleme cihazları, uzaktan algılama veya fabrika içerisinde taşımacılık hizmetini üstlenen robotik alanlarda büyük kullanım yelpazesine sahiptir. Açıklanabilir yapay zeka çalışmalarında piksel bazlı görüntüyü tasnif edebilme yeteneği sayesinde ilgi çekici bir problemidir. Gerçek zamanlı ve yüksek doğruluk gerektiren problemleri anlamsal bölütleme ile çözmek için çok sayıda çalışma vardır.

Anlamsal bölütleme problemlerin çözümünde derin öğrenme tabanlı modellerin kullanılması, klasik görüntü işleme algoritmaları tercih edilmemesi, görüntü işleme alanının daha çok yapay zeka tabanlı yöntemlere kaydığını göstermektedir. Bu döneme kadar görüntü sınıflandırma görevlerinde kullanılan geleneksel görüntü işleme yöntemlerin performanslarının sınırlarının olması, bu alanda çalışanların ve teknoloji geliştirenlerin daha iyi sonuç getiren çözümler bulma arayışına itmiştir. Anlamsal Bölütleme için Tümünden Evrimsel Ağlar (FCN for Semantic Segmentation) isimli çalışma, anlamsal bölütleme problemleri için önem arz etmektedir [7]. Bu çalışmada teklif edilen evrimsel sinir ağları, anlamsal bölütleme problemine klasik görüntü işleme algoritmalarından daha iyi performans vereceği gösterilmiştir. FCN yönteminin temel başarısı atlama bağlantılarıyla (skip connections) anlamsal boşluğu doldurmasıdır. Bu sayede model eğitimi sırasında bir anlam kaybı yaşanmayarak detaylar korunmuş olmaktadır.

FCN'nin ortaya koyduğu bu yaklaşımdan esinlenerek, birçok araştırmacı tarafından çok sayıda optimize yapay zeka modelleri ortaya konmuştur. Medikal görüntüleme alanında oldukça sık kullanılan U-Net mimarisi bunlardan biridir[8]. U-Net, simetrik bir kodlayıcı ile çözücü mimarisi ve atlama bağlantıları sayesinde az bir veri kümesi ile yüksek performanslı sonuçlar üretebilmektedir. U-Net'in sahip olduğu düşük veri ile yüksek performans özelliği, özellikle az bir görüntü veri kümesine sahip olan medikal görüntüleme alanında çok işe yaramaktadır.

DeepLab mimarisi ise daha genişli kapsamlı veri kümelerinde yüksek performanslı sonuç elde etmek üzere tasarlanmıştır[9]. DeepLab mimarisi, özellikle geniş bağlamsal anlam bilgisini yakalamak amacıyla geliştirilen atrous convolution ve Atrous Spatial Pyramid Pooling (ASPP) [10] yöntemlerini kullanarak farklı çözünürlük seviyelerinde etkili anlam çıkarımı yapabilmektedir. Uzaktan algılama, otonom sürüş ve deniz altı anlamlandırma gibi çalışmalarda yüksek başarımlarla sonuçlar üretebilmektedir. FCN[7], U-Net[8] ve DeepLab[9] gibi mimarilerin sağladığı bu önemli ileriye doğru atılan adımlar, anlamsal bölütleme alanının hızlı ve sürekli gelişimine katkı sağlamış ve daha sonra gelen birçok çalışmaya temel oluşturmuştur.

Derin öğrenme tabanlı yöntemlerinden elde edilen yüksek performanslı sonuçlar, eğitim sırasında kullanılan eğitime ait veri kümelerine aittir. Fakat gerçek dünyada farklı veri kümeleri üzerinde çalışacak bu modeller başka veri kümelerinin

dağılımlarını da büyük ölçüde anlamlandırabiliyor olması gerekmektedir. Gerçek dünyaya ait veri kümesi oluşturulmak istendiğinde, oluşturulabilecek olan veri kümesine ait görüntüler farklı coğrafi bölgelerden, farklı zaman dilimlerinden, farklı sensör tiplerinden ve hatta farklı hava koşullarından elde edildiğinde, eğitim ve test veri kümeleri arasında büyük dağılımsal farklılıklar (domain shift) ortaya çıkmaktadır. Bu dağılımsal farklılıklar, modelin eğitim aşamasında edindiği bilgi ve temsillerin, farklı veri kümelerinde geçerli olmamasına ve performansın ciddi oranda düşmesine neden olmaktadır.

Sentetik görüntüler, çoğunlukla video oyunları veya simülasyon programları gibi kullanıcı tarafından değiştirilebilir, kontrol edilebilir ortamlarda oluşturulabilmektedir. Bu görüntüler gerçek dünya görüntülerine göre daha az karmaşık yapılara sahiptirler. Özellikle toplanan bu sentetik görüntüler ile eğitilmiş modellerin gerçek dünya senaryolarında yetersiz performans gösterdiği gözükmemektedir[3]. Bu durumun temel nedeni, sentetik görüntülerin sahip olduğu farklı renk dağılımları, ışıklandırma koşulları, nesne dokuları ve genel görüntü stilinin gerçek dünya görüntüleriyle eşleşmemesidir. Bu farklılıklar, modelin eğitim sürecinde edindiği temsillerin yeni, daha önce hiç görmediği ortamlara genelleştirme yeteneğini önemli ölçüde sınırlar.

Ayrıca, gerçek dünya senaryolarında görüntülerin elde edildiği cihazların teknik özelliklerindeki farklılıklar ve görüntü kaydetme süreçlerindeki yöntemsel farklılıklar da alan kaymasını problemi sebep olmaktadır. Örneğin, farklı sensörlerin kullandığı görüntü çözünürlükleri, kontrast ve frekans alanındaki özellikler arasında önemli farklılıklar olabilir. Bu nedenle, bir veri setinde yüksek doğrulukla eğitilmiş bir model, farklı bir sensörle veya farklı bir ortamda alınan görüntüler üzerinde beklenen doğruluk seviyesine ulaşamayabilir. Bu durum, birçok gerçek dünya senaryolarının gerçekleştirilmesin modellerin geniş ölçekli uygulanabilirliğini engelleyen temel bir sorun olarak ortaya çıkmaktadır.

Son dönemde, alan uyarlama problemlerini çözmeye yönelik birbirinden farklı yöntemler ortaya çıkmıştır. Özellikle, görüntü stillerindeki farklılıkların frekans alanında daha etkili şekilde çözülebileceği gözlemi, araştırmaları bu yönde yoğunlaştırmıştır[3]. Bu yaklaşımın temel dayanağı, farklı alanlardan gelen görüntülerin stil farklılıklarının büyük oranda görüntünün düşük frekans

bileşenlerinde yoğunlaşmasıdır. Çünkü düşük frekans bileşenleri, genel olarak görüntünün aydınlatma, renk tonu ve doku gibi stil unsurlarını taşır ve yüksek frekans bileşenlerine kıyasla domain farklılıklarından daha fazla etkilenir.

Fourier alan uyarlaması ile kaynak alanın düşük frekanslı genlik bileşenleri, hedef alana ait bileşenler ile değiştirilir kaynak ve hedef alana ait görüntüler arasındaki dağılım farklılıkların azaltılması beklenir. Bu sayede kaynak alan görüntüleri, hedef alan görüntülerin stili ile daha uyumlu hale getirmektir. Bu tür yöntemler, görüntülerin temel stil özelliklerini kontrol ederek iki farklı alan arası tutarlılığı artırmakta ve böylece daha iyi model performansı yakalanmaktadır.

Ancak Fourier alan uyarlaması gibi temel yaklaşımlar, stil aktarımını basit sabit pencerelerle gerçekleştirdiğinden sınırlı bir stil aktarımı sunmaktadır. Bu sebeple, stil aktarımı yaparken hangi bileşenlerinin daha önemli olduğunu öne çıkarmak ve belirlemek için öğrenilebilir pencereci yaklaşımlar geliştirilmiştir. Bu tarz öğrenilebilir bir pencereye sahip yöntemler, belirli alanlarda veya belirli sınıflar için hangi frekans bileşenlerinin aktarılacağına otomatik olarak karar verebilmekte ve daha iyi alan uyarlama sonuçları elde etmektedir. Yapılan bu gelişmeler ışığında alan uyarlama problemlerine daha esnek ve etkin çözümler sunularak, modellerin farklı ve yeni alanlardaki uyarlama yeteneğini önemli ölçüde artırmıştır.

Bu tez çalışmasında ele alınan yöntemlerden biri olan frekans temelli optimal stil karışımı yönteminde [5], literatürdeki mevcut frekans tabanlı alan genelleştirme yöntemlerinden farklı olarak, stil bilgisini yalnızca gerekli durumlarda ve anlamsal bölütleme performansını olumsuz yönde etkilemeyecek şekilde aktarır. Geleneksel yöntemler genellikle stil bilgisinin tamamını ya da büyük bir kısmını doğrudan aktararak modelin orijinal anlamsal bilgileri kaybetmesine yol açabilirken, FOSMix yöntemi bu problemi özel olarak adresleyerek çözmektedir. Yöntem, frekans uzayında gerçekleştirilen stil aktarımını optimize etmek için sabit veya öğrenilebilir maske yapılarını kullanarak hangi frekans bileşenlerinin aktarılacağını dikkatli biçimde belirlemektedir. Bu maskeler, stil bilgisinin sadece anlamsal olarak kritik olmayan frekans bölgelerinde aktarılmasını sağlar ve böylelikle görüntülerin orijinal anlamsal özellikleri korunarak daha yüksek tutarlılık sağlanır. Ayrıca, öğrenilebilir maskeler ile model, veri setindeki özel alan özelliklerine göre kendini sürekli uyarlayabilir, bu da modelin genelleştirme yeteneğini artırır ve daha fazla alan çeşitliliğine dayanıklı hale

getirir. FOSMix, stil aktarımının sağladığı çeşitlilik ve alan farklarının giderilmesi ile anlamsal bölütleme modellerinin genelleştirme yeteneğini büyük ölçüde artıran yenilikçi ve etkin bir çözüm sunmaktadır.

Bu tez çalışması, anlamsal bölütleme alanında sıklıkla karşılaşılan ve gerçek dünya uygulamalarında kritik bir öneme sahip olan alan kayması problemini frekans tabanlı alan uyarlama yöntemleri ile çözüm bulmaya ve yöntemleri karşılaştırmaya çalışmaktadır. Çalışma kapsamında frekans tabanlı alan uyarlama yöntemleri incelenmiştir, bu yöntemlerin alan uyarlama problemlerindeki geçerliliği, farklı senaryolar ve uygulama alanları üzerinde karşılaştırmalı denemelerle analiz edilmiştir.

Yapılan denemeler ve bu denemelerin çıktıları incelenerek, Fourier Alan Uyarlaması (Fourier Domain Adaptation - FDA)[3], Gauss Fourier Alan Uyarlaması (Gaussian Fourier Domain Adaptation - GFDA)[4] ve farklı frekans maskeleye stratejileri içeren yöntemlerin avantajları ve dezavantajları verilmiştir. FDA yöntemi, temel Fourier dönüşümüne dayalı, hedef alan ile kaynak alan görüntüsüne ait genliklerin spektrumlarının bir pencere yardımıyla değiştirilmesi yaklaşımıyla alan uyarlama alanında göze çarpan bir yöntem olarak yer almıştır; ancak bu yöntem, sabit bir pencere sebebiyle, sabit frekans aralığı kullanmaktadır. Bu sebeple esneklik konusunda sınırlar içermektedir. GFDA yöntemi ise, bu sınırları azaltmak amacıyla düşük frekans bölgesini Gaussian filtresi ile daha yumuşak bir şekilde sınırlandırmaktadır. Ancak, bu yöntemlerin tümü, anlamsal bilgiye özel adaptasyon sağlayamamaları ve görüntülere ait stillerin anlamsal bilgi ile ilişkisiz bir şekilde aktarılması sebebiyle potansiyel anlamsal tutarsızlıklara yol açabilmektedir.

Literatürde yakın tarihte önerilen yöntemlerden biri Frequency-based Optimal Style Mix (FOSMix) yöntemidir[5]. Uygulanan bu yöntem, frekans tabanlı stil aktarımı tekniklerinin sağladığı avantajları daha da ileri götürmektedir. FOSMix yöntemi, frekans alanında optimal pencerelerle gerçekleştirilerek yalnızca anlamsal açıdan kritik olmayan frekans bölgelerinin aktarılmasını sağlamaktadır. Bu sayede, anlamsal boşluktaki tutarlılık korunurken, farklı alanlarda ortaya çıkan stil farklılıklarının etkisi azaltılmaktadır. Yapılan tez çalışması ile frekans tabanlı alan uyarlama teknikleri olan FDA, GFDA ve FOSMix yöntemleri karşılaştırmalı olarak incelenmiş ve çeşitli veri kümeleri ile farklı koşullar altında sistematik testler gerçekleştirilmiştir.

Yapılan bu tez çalışması, sentetik veriler ile gerçek dünya verileri arasında gerçekleştirilen frekans tabanlı alan uyarlama yöntemlerin pratik uygulama potansiyelini literatür verileriyle birlikte ortaya koymaktadır. Bu tez çalışması bahsedilen frekans taban alan uyarlama yöntemlerin gerçek dünya problemlerinde ortaya çıkan alan kayması problemlerine yönelik sunduğu çözümleri ayrıntılı bir şekilde analiz etmektedir. Frekans tabanlı stil aktarımı yöntemlerinin anlamsal bölütleme problem çözümlerinde alan kaymasını azaltması sebebi ile alan uyarlama yeteneğinin etkili olduğu saptanmıştır. Anlamsal bölütleme ve alan uyarlaması alanlarında gelecekte yapılacak olan araştırmalar ve gerçek hayatta uygulanacak projeler için geniş bir teorik temel ve pratik bir çerçeve sunmayı hedeflemektedir ve gelecekte yapılacak çalışmalara temel bir kaynak olma hedefi taşımaktadır.

Bu tez altı bölümden oluşmaktadır: Bölüm 2’de, anlamsal bölütlemenin akademide ve endüstriyel alanlardaki kullanım uygulamaları, yaygın veri kümeleri, performans metriklerini ve seçili çalışmaların mimari akış şemaları detaylı bir literatür taraması çerçevesinde sunulmaktadır; Bölüm 3’te, veri ön-işleme adımlarından başlayarak derin öğrenme model mimarisi ve önerilen frekans tabanlı alan uyarlama yaklaşımının ayrıntıları verilmektedir; Bölüm 4, kaynak alan veri kümesi ve hedef alan veri kümelerinin tanıtımını, farklı modelleme yöntemleri ile yapılan çalışmaları ve bu çalışmaların literatürdeki sonuçlarla karşılaştırılmasını içermektedir; Bölüm 5’te, elde edilen bulgular bütüncül olarak tartışılarak geliştirilen yöntemlerin hedeflere hangi ölçüde ulaştığı ve yöntemler arasındaki performans farklarının olası nedenleri değerlendirilmektedir; son olarak Bölüm 6, çalışmanın temel katkılarını özetlemekte ve gelecekte yapılabilecek geliştirmeler için öneriler sunmaktadır, böylece okuyucu araştırmanın adımlarını baştan sona tek bir akış içinde takip edebilmektedir.

2. LİTERATÜR TARAMASI

2.1. Derin Öğrenme

Derin öğrenme, giriş ile çıkış katmanları arasında çok sayıda bağlantı içeren yapay sinir ağlarının eğitiminde kullanılan yöntemler topluluğudur [11]. Derin öğrenmede öğrenilmesi gereken parametre sayısı geleneksel yöntemlere kıyasla çok daha fazladır. Bu yüzden bu tarz bir yöntemle derin anlamı verilmiştir. Derin öğrenmede temel olarak 3 farklı katman vardır. Giriş, çıkış ve gizli katmanlar derin öğrenme tabanlı modelleri özetleyebilir. Birçok görüntü işleme probleminde artık yüksek teknoloji ile yüksek performanslı bir şekilde kompleks eğitimleri mümkün kılındığından dolayı klasik görüntü işleme yöntemleri yerine artık derin öğrenme tabanlı yöntemler tercih edilmektedir.

Evrişimsel sinir ağlarında gizli katmanlar, genellikle konvolüsyon ve havuzlama (seyrek örnekleme) katmanlardır. Her evrişim katmanında, küçük boyutta bir filtreyi görüntü boyunca hareket ettirerek konvolüsyon operasyonu yapılır. Konvolüsyon operasyonunda filtredeki ve görüntüdeki piksellerin matriks elementleri çarpılır ve bulunan değerler toplanır.

Filtrenin değerleri iteratif eğitim süreci ile ayarlanır ve nöral ağ belli sayıdaki aşamalar için eğitildikten sonra filtreler artık görüntüde daha çeşitli özellikler için tetikte olur. Ağın ilk katmanları uçları tespit etmek gibi basit özellikleri arar. Ağda ilerledikçe daha kompleks özellikler tespit edilmeye çalışılır. Yani özellikler düşük seviyeliden yüksek seviyeliye çıkar. Havuzlama katmanları, görüntüyü seyrek örnekleme için kullanılır. Görüntü birçok piksel değeri barındırır ve tipik olarak ilerledikçe görüntü boyutu azalırsa ağın özellikleri öğrenmesi kolaylaşır. Havuzlama katmanları gerekli parametrelerin sayısını azaltır dolayısıyla hesaplamalar da azalır[12]. Havuzlama aynı zamanda aşırı öğrenmeden kaçınmayı sağlar. Farklı havuzlama katmanları bulunmaktadır.

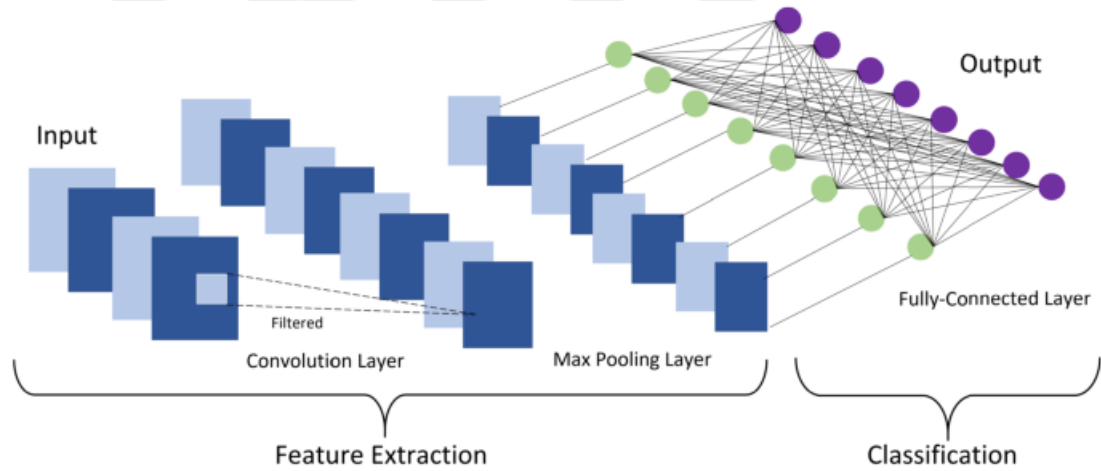
CNN'ler, görüntüden belli özellikleri sınıflandıran ve görsel analizde genişçe kullanılan bir derin öğrenme ağıdır. Uygulamaları, görüntü ve video tanımadan görüntü tasnifine; medikal görüntü analizinden doğal dil işlemeye kadar uzanır.

“Konvolüsyon” terimi matematiksel konvolüsyonu işaret eder. Bir fonksiyonun biçiminin diğer fonksiyon tarafından nasıl modifiye edildiğini gösteren üçüncü bir fonksiyonu elde etmek için 2 fonksiyonun çarpıldığı lineer bir operasyondur. Basitçe, matrisle gösterilen 2 görüntü, görüntüden özellik çıkarmı yapmak için çarpılır.

Evrişimsel sinir ağı mimarileri iki ana gövdeden oluşmaktadır. Bunlar:

1. Özellik çıkarmı (Feature extraction) denilen süreçte analiz için görüntünün çeşitli özniteliklerini tespit edilir ve birbirinden ayrıştırılır.
2. Tam bağlı (Fully connected) denilen, evrişim sürecinin çıktısından faydalanan ve önceki aşamada çıkarılmış görüntüye ait özellikleri temel alarak görüntünün ait olduğu sınıfı tahmin eden katmandır.

CNN Mimarisine Şekil 2.1 de yer verilmiştir.



Şekil 2.1: Örnek evrişimsel sinir ağı [13].

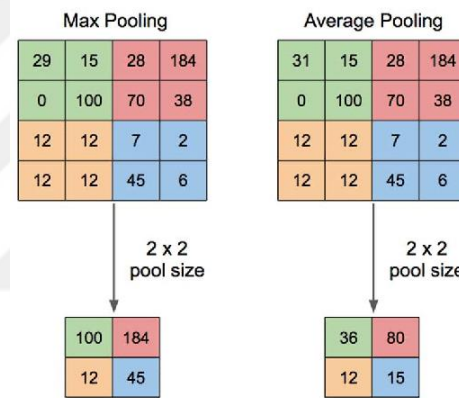
Evrişimsel sinir ağlarını oluşturan temel üç katman vardır. Bunlar sırasıyla: evrişimsel katmanları, havuzlama katmanları ve tam bağlı sinir ağı katmanlarıdır. Bu katmanlar birleştirildiğinde CNN mimarisi ortaya çıkmaktadır. Bu üç katmana ek olarak dropout [5].ve aktivasyon fonksiyonları da CNN mimarisi içerisinde önem arz ederler [5].

Evrişimsel katmanlarda, girdi görüntüsüne ait çeşitli özellikleri çıkaran model mimarisinin ilk katmanıdır. Bu katmanda, $M \times M$ boyutlu filtre ile girdi görüntüsü arasında konvolüsyon işlemi yapılır. Konvolüsyon işleminde kullanılan filtre, görüntü üzerinde kaydırılarak skaler çarpım işlemi uygulanır. Bu işlemin sonucunda ortaya çıkan çıktıya özellik haritası (feature map) denmektedir. Özellik haritaları görüntüde

bulunan detayla ilgili bilgi vermektedir. Elde edilen özellik haritaları kullanılarak sonraki katmanlar beslenerek görüntüde bulunan öznitelikler ile eğitim iyileştirilmeye çalışılır.

Çoğu problemde evrimsel katmanları, havuzlama katmanı takip etmektedir. Bu katmanın birincil amacı konvolüsyona uğramış özellik haritasının boyutunu azaltarak hesaplama işlemini hafifletmesidir. Her özellik haritasına bağımsız şekilde havuzlama uygulanır. Farklı havuzlama katmanları bulunmaktadır. Bunlardan ikisi:

- Maksimum Havuzlama (Max Pooling): Havuzlanan alandan en büyük değer seçilir.
- Ortalama Havuzlama (Average Pooling): Havuzlanan alanda bulunan değerlerin ortalaması alınır.



Şekil 2.2: Maksimum havuzlama ve ortalama havuzlama [12].

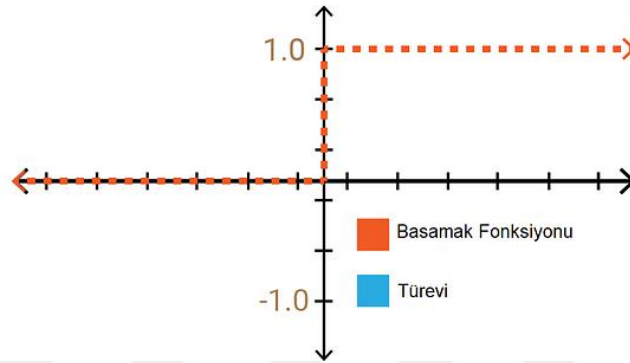
Havuzlama katmanları genellikle evrimsel katman ile tümünden bağlı sinir ağı katmanları arasında köprü görevi görmektedir.

Tümünden bağlı katman, ağırlıklardan (weights) ve eğri uzaklıklarından (bias) oluşmaktadır. Farklı katmanlardaki sinirleri birbirine bağlamak için kullanılmaktadır. Bu katmanlar genelde çıktıyı almadan önce CNN mimarisinin son kısmında bulunan katmanları oluşturur. Önceki katmanlardan gelen girdiler düzleştirme işlemine sokulur ve tümünden bağlı katman beslenir. Düzleştirilmiş vektör, matematiksel operasyonların yapıldığı katmanlardan geçerek tahminde bulunulur.

Özellik çıkartıcıdan çıkan çıktıların hepsi tümünden bağlı katmanlara girdi olarak girdiği zaman hem işlem gücü hemde yaklaştırılması gereken ağırlık sayısı model için artış göstermektedir [13]. Böyle bir mimari ile aşırı öğrenmeye yönelim gösteren bir

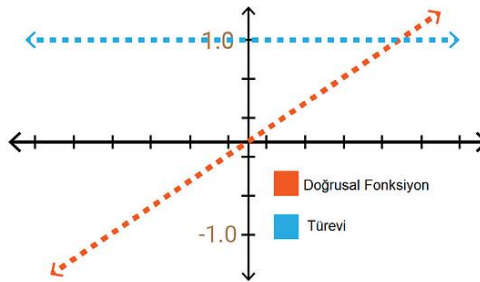
modelden kaçınmak için atma (dropout) işlemi uygulanır. Eğitim sırasında bazı sinir ağırları devre dışı bırakılarak eğitime o şekilde devam edilir. Devre dışı bırakılan sinir uçları her bir epoch boyunca değişiklik gösterebilir. Bu da modelin çeşitli bilgi farklılıkları ile öğrenerek aşırı öğrenmeden kaçınmasını sağlamaktadır.

Aktivasyon fonksiyonları her bir katmandan sonra model mimarisine eklenerek katmanların çıktısının etki eden matematiksel birer fonksiyondur. Literatürde çokça kullanılan basamak fonksiyonlarına aşağıda yer verilmiştir. [14]



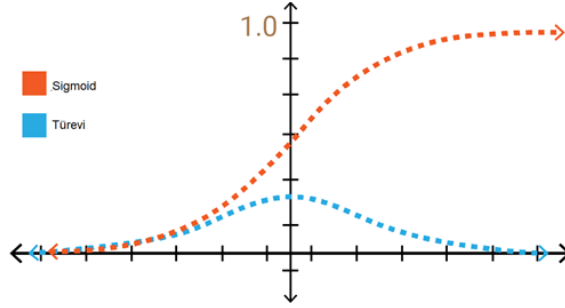
Şekil 2.3: Basamak fonksiyonu [14].

Basamak fonksiyonu, iki sınıfı ayırmak için kullanılır. Bu nedenle genellikle sinir ağlarının çıkış katmanında yer almaktadır. Ancak gizli katmanlarda kullanılması önerilmemektedir çünkü türevleri öğrenme sürecine katkı sağlayamamaktadır.



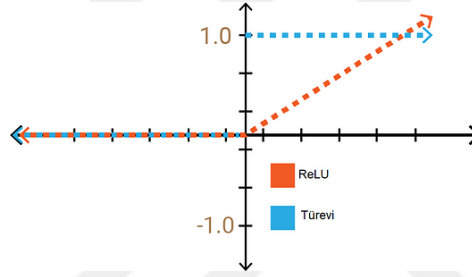
Şekil 2.4: Doğrusal fonksiyon [14].

Doğrusal fonksiyon, çıktı olarak ikili değerler yerine farklı farklı aktivasyon seviyeleri üretmektedir. Böyle bir yaklaşım ile sinirler arasında bağlantı kurulmuş olur. Fakat bu fonksiyonun türevi sabittir. Eğer modelin bütün katmanlarında bu fonksiyon kullanılmak istenirse, girişten çıkışa kadar her seviyede aynı doğrusal ilişki kurulmuş olur. Böyle bir yaklaşım ile öğrenme işlemi zorlaşmış olur.



Şekil 2.5: Sigmoid fonksiyonu [14].

Basamak fonksiyonuna benzer bir yapıya sahiptir fakat türevlenebilmektedir. Katmanlardan sonra kullanılabilir gradyanlar güncellenerek öğrenilebilir bir model oluşturulabilir.



Şekil 2.6: ReLU Fonksiyonu [14].

ReLU fonksiyonu, sinir ağlarında sıkça kullanılan bir aktivasyon fonksiyonudur. Pozitif girişleri direkt geçirmektedir. Negatif girişleri ise sıfıra çekmektedir. ReLU fonksiyonu kullanıldığında katman çıkışlarında sadece sıfır değeri ile sonsuz arasındaki değerler elde edilmektedir. Sigmoid fonksiyonun ürettiği negatif değerleri üretmediği için daha az işlem gücü gerektirmez. Bu sayede model eğitimi daha hızlı bir şekilde eğitilebilir. Her şeyin bir avantajı ve dezavantajı olduğu gibi ReLU fonksiyonunda vardır. Girdi yapılan değerler sıfırdan küçük değerliyse, ReLU fonksiyon çıktısı her zaman sıfır değerini alır. Bu durum öğrenmenin durmasına neden olabilir. Bu soruna "ölü sinir ucu" problemi denir.

2.2. Anlamsal Bölütleme

Anlamsal bölütleme (AB), bilgisayarla görü alanında sahneyi anlamlandırmaya yönelik temel ve yapılması zor görevlerden biridir. Anlamsal bölütleme yönteminde amaç, bir görüntüde yer alan her pikselin ait olduğu anlamsal sınıfların (örneğin, yaya,

yol, araç, trafik işareti vb.) belirlenmesidir. Bu yönüyle, yalnızca objenin varlığını veya konumunu değil; aynı zamanda tüm sahnenin piksel düzeyinde ayrıntılı sınıfsal bir piksel haritasını çıkarmayı hedefler. AB çıktısı sanki bir siyasi haritaya benzer boyanan bölgeler sınırlara sahiptir. Sınırlı alan içerisinde kalan kısımlar bir anlam ifade ederler. Anlamsal bölütleme uygulamaları otonom sürüş[2], yapı inşaatı alanlarında[15], insan-robot etkileşimi ve sanal gerçeklik [16] olmak üzere uygulama alanlarında çokça kullanılmaktadırlar. Hastane, tıp merkezleri gibi sağlık kuruluşlarında yer alan medikal görüntüleme cihazlarında da anlamsal bölütleme kullanılmaktadır[8]. Hastalık bulunan bölgelerin piksel düzeyinde tespiti gibi yüksek doğrulukla tahmin yeteneği gerektiren işlerde yer almaktadır.

Anlamsal bölütleme, çoğu zaman obje tespiti ile karıştırılmaktadır. Ancak iki problem birbirlerinden oldukça farklıdır. Obje tespiti, bir görüntüdeki nesneleri sınırlayıcı kutularla ile sınıflandırırken, anlamsal bölütleme görüntüde bulunan her bir piksele ilgili sınıfı atamaya yönelik çalışmaktadır. Bu sayede görüntünün her bir pikseli detaylandırılmış olur. Örnek vermek gerekirse, birbiri ile dip dibe veya çakışan objeler, obje tespiti yöntemlerinde ayrıştırılamazken, anlamsal bölütleme bu ayrımı net biçimde yapabilir [17]. Buna ek olarak, durum bölütlemesi (instance segmentation) yöntemleri (Mask R-CNN[18]), aynı sınıfa ait farklı objeleri birbirinden ayırarak her bir objeye ait piksel haritasını ayrı ayrı üretir.

Anlamsal bölütleme, ilk yıllarında öznitelik çıkarım yöntemlerine dayanmaktaydı. SIFT (Scale-Invariant Feature Transform) [19] ve HOG (Histogram of Oriented Gradients)[20] yöntemleri ile düşük seviyeli öznitelikler çıkartılarak görüntü anlamlandırılmaya çalışılıyordu. Ancak bu yöntemler, özellikle karmaşık görüntülerde düşük bir performans göstermekteydi. Son yıllarda derin öğrenme yaklaşımları bu alanda daha yüksek başarımlar elde ettiğinden dolayı uygulama alanında daha çok kullanılmaya başlanmıştır.

2.2.1. Anlamsal Bölütleme Yaklaşımları

2.2.1.1. Bölge Temelli Anlamsal Bölütleme:

Bölge temelli metotlar genelde tanıma kullanarak bölütleme yapar.

Bu yöntemde, öncelikle görüntüden belli bölgeler çıkartılır ve onları tasnif edilir. Sonraki aşamada bölge temelli tasnif yapılır. Test sürecinde bölge temelli tahminler piksel tahminlerine dönüştürülür. Genelde pikseli içeren bölgeler arasında en yüksek puanı olan ile etiketlenir. R-CNN (Regions with CNN feature)[21], bölge temelli metotlardan biridir. Obje tespitindeki sonuçlara dayanarak anlamsal bölütleme yapılmaktadır. R-CNN, seçici bir arama ile önce büyük miktardaki obje önerilerini çıkarır ve her birinin CNN özelliklerini hesaplar. Son olarak her bölgeyi sınıfa özgü lineer SVM'ler[22] ile sınıflandırır. Geleneksel CNN yapılarının başlıca amacı görüntü tasnifi iken R-CNN[21], nesne tespiti ve görüntü bölütleme gibi daha karmaşık görevleri başarabilir ve her iki alanda da önemli bir temel haline gelir. Dahası R-CNN, AlexNet[23], VGG[24], GoogLeNet[25] ve ResNet[26] gibi herhangi bir CNN yapısının üstüne inşa edilebilmektedir. Görüntü bölütleme görevi için R-CNN, her bölge için iki tip özellik çıkarır. Tam bölge özelliği (full region feature) ve ön plan özelliği (foreground feature). Bu iki özelliği ardarda eklenerek 1 bölge özelliği elde edilerek daha iyi bir performans alınacağı keşfedilmiştir. R-CNN, büyük ölçüde ayırt edilebilen CNN özellikleri kullanabildiğinden iyi bir performans göstermektedir. Ancak bölütleme görevi için birkaç engelden dolayı da kötü etkilenir. Özellikler, anlamsal bölütleme görevi ile uyumlu olmayabilir. Özellikte isabetli bir sınır üretimi için yeterli uzaysal bilgi eksiktir. Bölüt temelli öneriler üretmek zaman alan bir iş ve son performansı çok etkiler.

2.2.1.2. Tümünden Evrişimsel Ağ Temelli Anlamsal Bölütleme:

FCN (tümünden evrişimsel ağ/fully convolutional network) pikseller arasındaki eşlemeyi bölge önerilerini çıkarmadan öğrenir[7]. FCN ağı, klasik CNN'in uzantısıdır. Burada ana fikir, FCN'in herhangi bir boyuttaki görüntüleri girdi olarak almasıdır. CNN'lerdeki kısıtlayıcı faktör, sadece belli boyuttaki girdileri kabul edip etiket üretmesidir. Bunun sebebi tam bağı katman sayısının sabit olmasıdır. Bunun aksine FCN'ler sadece konvolüsyonel ve havuzlama katmanlarına sahiptir. Bu sayede herhangi bir boyuttaki görüntüleri girdi kabul edip tahminde bulunabilmektedirler. FCN'lerdeki bir problem, değiştirilmiş birkaç evrişimsel katmanlardan ve havuzlama katmanlarından geçildiği için çıktındaki özellik haritası seyrek örneklemeye uğramaktadır. Bu yüzden FCN'deki doğrudan yapılan tahminler tipik olarak düşük çözünürlüklüdür ve bu yüzden obje sınırları bulanık olmaktadır. Bu sorunu gidermek

için çeşitli FCN modeli temel alınarak daha ileri seviye yaklaşımlar teklif edilmiştir. Bu yöntemlerden bazıları SegNet[27], DeepLab[9] ve U-Net[28] gibi daha karmaşık yapıya sahip modellerdir.

2.2.1.3. Az Denetlenen Anlamsal Bölütleme:

Anlamsal bölütleme ile ilgili metotların çoğu piksel bölütleme maskeleri olan büyük sayıdaki görüntülere dayanır. Ancak manuel olarak bu maskeleri ifade etmek çok zaman alır ve ticari olarak pahalıdır. Bu yüzden son zamanlarda bazı az denetlenen metotlar teklif edilir. Bu metotlar, sınırlı kutulardan faydalanan anlamsal bölütlemeyi gerçekleştirmeyi amaçlamaktadır. [29]

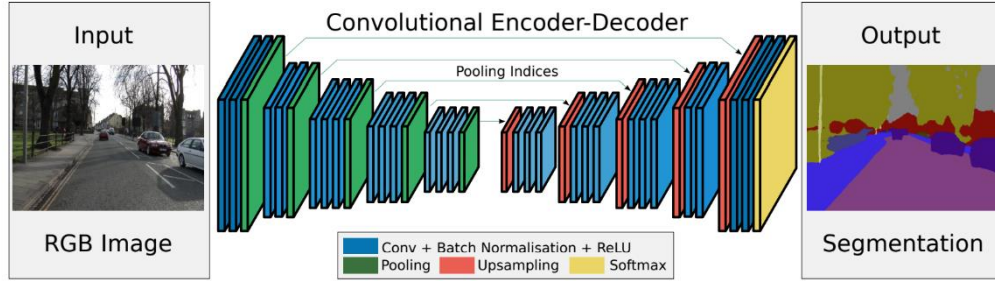
2.2.2. Anlamsal Bölütlemeye Kullanılan Modeller ve Açıklamaları

Tipik bir anlamsal bölütleme mimarisi, kodlayıcı (encoder) ağını takip eden bir çözücü (decoder) ağı olarak düşünülebilir. Kodlayıcı genelde önceden eğitilmiş VGG16[24], ResNet-18 ve ResNet-50[26] gibi tasnif ağlarıdır (backbone). Bu ağı çözücü ağı takip etmektedir. Çözücünün görevi, yoğun tasnif elde etmek için kodlayıcının yüksek çözünürlüklü piksel uzayında öğrendiği düşük çözünürlüklü ayırıcı özellikleri (discriminative features) anlamsal olarak göstermektir. Tasnifte sadece çıktıdaki sonuçlar önem arz ederken, anlamsal bölütlemeye piksel düzeyinde ayıklamanın yanında kodlayıcının piksel uzayında farklı aşamalarda öğrendiği ayıklayıcı özellikleri göstereceği bir mekanizma da gereklidir. Farklı yaklaşımlar çözücü mekanizması için farklı mekanizmalar benimsemektedir.

2.2.2.1. SegNet Mimarisi

SegNet mimarisi, anlamsal bölütleme problemleri için geliştirilen bir modeldir.[27] Şekil 2.7’de model mimari incelenebilir. Temel olarak iki farklı kısımdan oluşur. Bunlar kodlayıcı (encoder) ve çözücü (decoder) yapısına sahiptir. İlk kısmı yani kodlayıcı kısmı, görüntüden anlamsal bilgi çıkarımı yaparken, çözümleyici kısmı ise konumsal detayları geri kazanmak amacıyla model mimarisinde yer alır. Kayıplı maksimize havuzlama (max pooling) işlemleri sırasında tutulan konum indislerini geri kazandırmak amacıyla çözücü kısmı kullanılmaktadır. Bu sayede bölütleme çıkışıdaki sınırlar daha keskin ve lokal detaylar belirgin bir hale getirilmektedir.

Sınırlı sayıda parametre ile yüksek doğruluklu bir model üretimi nedeniyle klasik CNN temelli modeller arasında SegNet mimarisine sahip modeller dikkat çekmektedir.

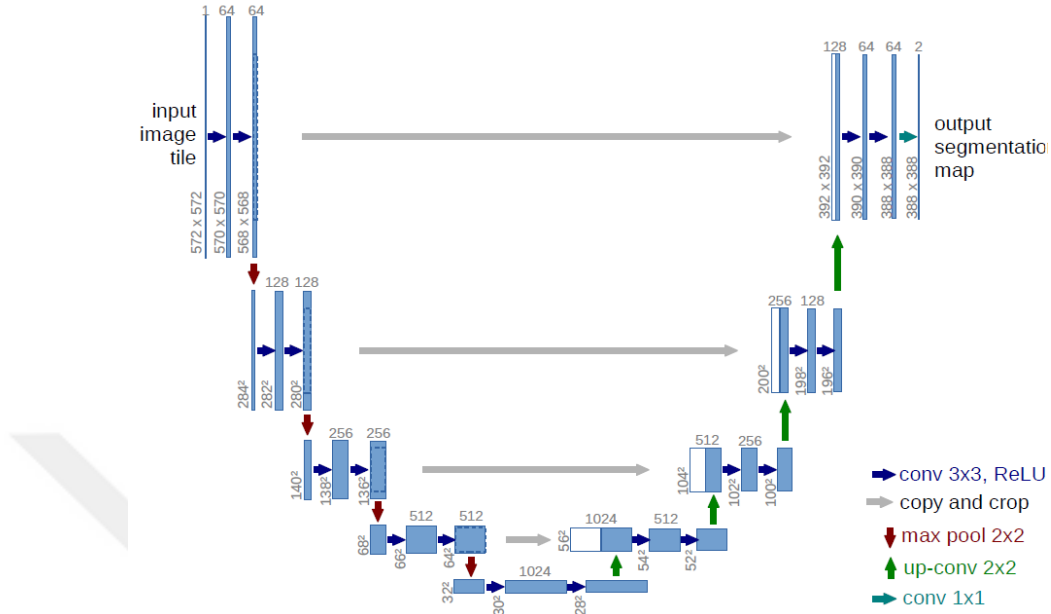


Şekil 2.7: SegNet mimarisi [27].

2.2.2.2. U-Net Mimarisi

U-Net, görüntülerin hızlı ve titiz bölütlenmesi için tasarlanmış evrişimsel sinir ağı mimarisidir[28]. Bugüne kadar birkaç zorlu problemlerde gösterdiği iyi performans bu modeli öne çıkarmıştır. Anlamsal bölütleme alanında en popüler mimarilerden birisidir. Bu ağ önceden bahsedildiği gibi yine kodlayıcı kısmı ve çözücü kısmından oluşmaktadır. Kodlayıcı, görüntülerin farklı ölçeklerdeki özelliklerini geleneksel evrişimsel ve havuzlama katman yığınları ile yakalar. Somut olarak, kodlayıcı bloğunda tekrarlı bir şekilde kullanılan iki tane konvolüsyonel katman, onu takip eden havuzlama katmanı vardır. Her evrişimsel blok ve ona bağlı havuzlama operasyonu için, ağın karmaşık yapıları verimli bir şekilde öğrendiğinden emin olmak amacıyla özellik haritası sayısı iki katına çıkarılmaktadır. Çözücü kısmı simetrik şekilde genişleyen ve kodlayıcıya benzeyen bir kısımdır. Transpozu alınmış konvolüsyonlar kullanır. Bu tip evrişimsel katman, eğitilebilir parametreleri olan sık örnekleme metodudur ve maksimum havuzlama gibi katmanların tersini yapmaktadır. Kodlayıcıya benzer şekilde her konvolüsyon katmanı sık evrişimsel katmanı (up-convolutional layer) takip eder. Her blokta özellik harita sayısı yarıya iner. Küçük özellik haritasından bölütleme maskesini tekrar oluşturmak zor bir görev olduğundan her sık evrişimsel katmanın çıktısına, ona karşılık gelen kodlayıcının özellik haritası eklenir. Kodlayıcı katmanının özellik haritası, eğer boyutlardan biri karşılık gelen çözücü katmanını aşarsa kırpılır. Son olarak çıktı, özellik harita sayısının tanımlı etiket sayısına eşit olduğu başka bir evrişimsel katmandan geçer. Şekil 2.7’de model

mimarisi incelenebilir. Bu mimariler, görüntüdeki düşük seviyeli öznelikleri yüksek seviyeli anlamsal bilgiyle birleştirerek daha hassas sınır tespiti sağlar.



Şekil 2.8: U-net mimarisi [28].

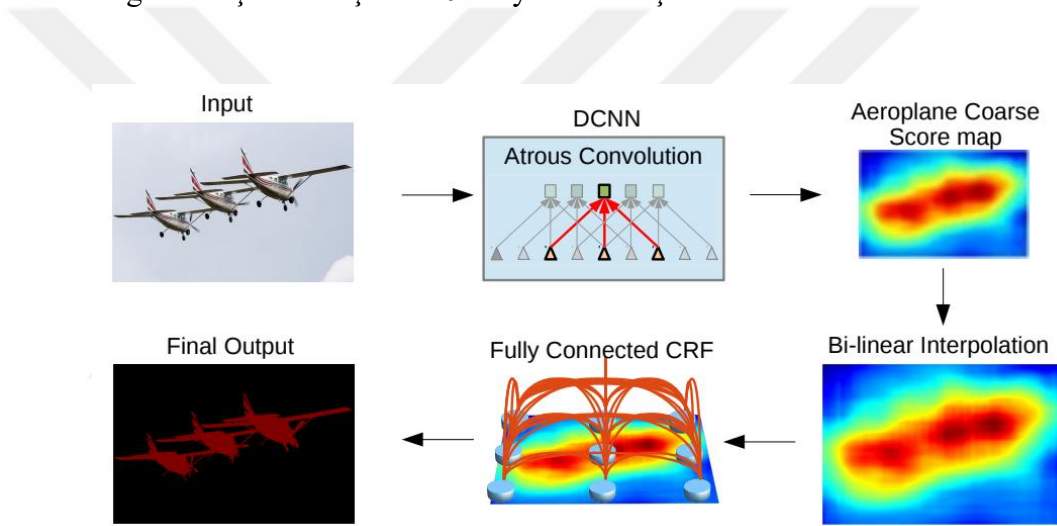
2.2.2.3. DeepLab Mimarisi

DeepLab[9] mimarisi evrimsel sinir ağı temelli state-of-the-art bir anlamsal bölütleme mimarisidir ve güncel çalışmalarda genel olarak yüksek performans sergilemektedir. Özellikle görüntülerdeki kompleks yapıları doğru şekilde sınıflandırmak için tasarlanmış olup özellikle farklı boyut ve konumlardaki objelerin etkin bir şekilde modellenmesini sağlayacak konum bilgisini işleme yeteneğine sahiptir. DeepLabv2 ise bu serinin önemli ikinci temsilcisi olarak esnek yapısıyla çoklu özel bilgiyi yakalama yeteneği ile ön plan çıkar.

DeepLab versiyon 2 mimarisinin ana bileşeni genişletilmiş evrimsel (dilated convolution) katmanları olarak kabul edilir. Geleneksel evrim filtrelerinin etkinliğini artırmak için özelliklere boşluklar ekleyerek modelin görüntüleri daha geniş kapsamda tanıması yeteneğini arttırmaktadırlar. Bu sayede alıcı alan genişletilirken çözünürlük kaybolmadığından görüntüde bulunan ayrıntılar korunmaktadır. Geniş kapsamlı ilişkilere odaklanılır. Bu özellik özellikle ince kenarlı nesne veya yakın

konumlandırılmış küçük nesne grupları gibi hassas ayırım gerektiren durumlarda önemli avantaj sağlar.

DeepLabV2'nin önemli bileşenlerinden birisi Atrous Uzaysal Piramit Havuzu (Atrous Spatial Pyramid Pooling -ASPP)[10] yapısını barındırmasıdır. Birden fazla genişleme oranı ile tanımlanmış paralel evrişimsel katmanlarından oluşmaktadır. Her bir katmanın amacı ayrı ayrı anlamsal bilgi düzeyleri yakalamaktır. Örneğin, düşük genişletme oranları yerel detayları yakalarken yüksek oranlar daha geniş açıları algılayabilir. Bu paralel yapının birleştirilmesi sayesinde her piksel için hem yerel hem de küresel düzeyde zengin temsillere ulaşıyor. Böylece hem küçük nesneler hem de geniş yapılar aynı anda bölütleme sürecine tabii tutulmaktadır. Örnek bir DeepLab model girdi ve çıktısına Şekil 2.9'da yer verilmiştir.



Şekil 2.9: DeepLab örnek akış şeması.

Kodlayıcı kısmı, genellikle VGG16[24] veya ResNet[26] gibi çok kullanılan bir backbone ile kurulmaktadır. Bu sinir ağı, görüntüden yüksek seviyeli anlamsal özellikleri çıkararak, anlamsal bölütleme görevine temel oluşturmaktadır. Ancak klasik çözümleyici kısmı, U-Net'teki [28] gibi açıkça tanımlanmış bir yukarı örnekleme mimarisi içermemektedir. Düşük çözünürlüklü çıktı, lineer olmayan interpolasyon gibi yöntemlerle orijinal çözünürlüğe görüntüyü geri ölçeklemektedir. Tasarımın bu şekli oluşturulan modelin bilgisayar hafızasında daha verimli çalışmasını sağlarken, bölütleme doğruluğunu da makul düzeyde tutar.

Sonuç olarak, DeepLabv2 mimarisi, özellikle kompleks görüntülerin veya çok sınıflı anlamsal bölütleme problemlerinin çözülmesinde fazla bir detay kaybı yaşamadan

geniş örüntü yakalayabilen, hesaplama açısından aşırı parametresi olmayan, dengeli ve etkili bir çözüm sunmaktadır. ASPP ile sağlanan çoklu ölçekli temsil yeteneği ve genişletilmiş konvolüsyonlar sayesinde bu model, anlamsal bölütleme literatüründe önemli bir yeri vardır. Güncel çalışmalarda çokça kullanılmaktadır.

2.2.3. Anlamsal Bölütlemeye Kullanılan Metrikler

2.2.3.1. Mean Intersection Over Union

Anlamsal bölütlemeye metrik değer ölçümü için bir önceden etiketlenmiş maskeler (ground-truth) ile bu maskelere eşlik eden görüntülerin girdi yapılarak modelin tahmin ettiği çıktı maskeleri karşılaştırılarak metrik hesabı yapılır. Bu çıktı sonucunda örtüşme var mı yok mu diye kontrol ederek her bir piksel değeri 0 ile 1 arasında değer alır. 1 değeri tamamen aynı olduğunu söyler; sıfır değeri ise tamamen farklı olduğunu söyler. Veri kümesinde bulunan bütün sınıflar için birleşimlerin kesişimi (Intersection over Union - IoU) değerleri hesaplanır ve bunun ortalaması alınırsa elde edilen metriğe ortalama birleşimlerin kesimi (mean Intersection over Union mIoU) denmektedir.

$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (2.1)$$

- A: Model Tahmini
- B: Ground-Truth Maske

$$mIoU = \frac{1}{n} \sum_{i=1}^n IoU \quad (2.2)$$

- n: Toplam Sınıf Sayısı

2.2.3.2. Piksel Doğruluğu (Pixel Accuracy)

Doğruluk puanı, algoritmanın doğru tahminleri ile diğer yaptığı bütün tahminlere oranıdır.

$$\text{Doğruluk} = \frac{\text{Doğru Pozitif} + \text{Doğru Negatif}}{\text{Doğru Pozitif} + \text{Yanlış Negatif} + \text{Doğru Negatif} + \text{Yanlış Pozitif}} \quad (2.3)$$

- y: Model Tahmini
- Z: Ground-Truth Maske
- n: Örnek Sayısı

2.2.3.3. Keskinlik (Precision)

$$\text{Keskinlik} = \frac{\text{Doğru Pozitif}}{\text{Doğru Pozitif} + \text{Yanlış Pozitif}} \quad (2.4)$$

2.2.3.4. Duyarlılık (Recall)

$$\text{Duyarlılık} = \frac{\text{Doğru Pozitif}}{\text{Doğru Pozitif} + \text{Yanlış Negatif}} \quad (2.5)$$

2.2.3.5. F1 Skoru

$$\text{F1 Skoru} = 2 * \frac{\text{Keskinlik} * \text{Duyarlılık}}{\text{Keskinlik} + \text{Duyarlılık}} \quad (2.6)$$

2.2.3.6. Dice Skoru

$$\text{Dice Skoru} = 2 * \frac{\text{Doğru Pozitif}}{(\text{Doğru Pozitif} + \text{Yanlış Pozitif}) + (\text{Doğru Negatif} + \text{Yanlış Pozitif})} \quad (2.7)$$

2.3. Alan Uyarlama

Anlamsal bölütleme de kullanılan derin öğrenme tabanlı modeller, yüksek doğrulukta sonuç elde edebilmek için büyük ve tekdüze dağılıma sahip veri kümelerine ihtiyaç duyarlar. Gerçek dünya üzerinde böyle bir veri oluşturmak mümkün olmayabilir. Örnek vermek gerekirse, araba ile kaydedilen bir veri kümesinin büyük bir çoğunluğunu yol, binalar veya araçlar yer alacaktır. İnsan, hayvan veya bisiklet gibi

farklı diğer sınıfların dağılımı düşük olacaktır. Anlamsal bölütleme için bu tarz bir veri kümesini topladıktan sonra belli sınıflara göre etiketleme işlemine tabi tutulup her bir görüntüye karşılık gelen bir piksel haritası oluşturmak gerekecektir. Bu da çok zaman alıcı ve maliyetli bir işlemdir. Farklı bir sorun ise toplanan verinin alan bağımsız olmamasıdır. Örneğin eğitim yapılacak veri ile test yapılacak veri birbirinden konumu, hava durumu ve mevsim gibi birçok parametre tarafından farklılık gösterebilir. Bu sebeple eğitilen modelin performansı değerlendirilirken düşük performans elde etmek kaçınılmaz olabilir. Bu dağılımsal farklılığa alan kayması (domain shift) adı verilir.

Alan kayması problemini gidermek için yapay zeka alanında, alan uyarlama (domain adaptation) tekniklerinden yararlanılmaktadır. Bu sayede ortamlara ait dağılımsal farklar saptanıp, ortadan kaldırılabilir veya fark azaltılabilir. Alan uyarlamasında iki farklı alan vardır bunlar sırasıyla: kaynak alan (source domain) ve hedef alandır (target domain). Kaynak alan, yapay zeka modelinin eğitimin yapılacağı veri kümesine ait alandır. Hedef alan ise test edilmesi planlanan alandır. Amaç, kaynak alan ile hedef alan veri kümeleri arasında benzerlik kurarak bir bilgi köprüsü oluşturmaktır. Oluşturulan bu bilgi köprüsü ile kaynak alanda eğitimi gerçekleştirilen model performansının, hedef alanı kapsayan veri kümesi üzerinde iyi bir performans vermesi beklenmektedir. Alan uyarlama alanında farklı senaryolar ile farklı alan uyarlama yöntemleri oluşturulabilmektedir.

2.3.1. Alan Uyarlama Yöntemleri

Alan uyarlama yöntemleri üç gruba ayrılmaktadır:

2.3.1.1. Denetimli Alan Uyarlama:

Denetimli alan uyarlamasında hem kaynak alan hem de hedef alanda her bir veriye karşılık gelen etiketli veri bulunmaktadır. Bu tarz problemlerde genellikle veri sınırlı yani az olmaktadır.

2.3.1.2. Yarı Denetimli Alan Uyarlama

Yarı denetimli alan uyarlaması, hedef alana ait hem etiketli hem de etiketsiz verilere sahip olduğunda danışılan bir yöntemdir[30]. Bu kompleks eğitim yöntemi genel olarak etiketsiz veriden öğrenmeye modeli teşvik etmektedir.

2.3.1.3. Denetimsiz Alan Uyarlama

Bu tarz bir alan uyarlaması diğer alan uyarlama göre en zorlayıcı alan uyarlaması tekniğidir [31]. Modelin öğrenebileceği hedef alanda etiketli bir veri bulunmamaktadır. Model sadece kaynak alana ait verileri kullanarak eğitilerek, hedef alanda test edilir. Bu yaklaşım gerçek dünya senaryolarını tamamen kapsamaktadır. Eğer ki etiketsiz bir veri üzerinde iyi bir performans elde ediliyorsa gerçek uygulamalarda rahatlıkla eğitilen modeller kullanılabilir.

2.3.2. Özellik Seviyesi Alan Uyarlama

Kaynak ve hedef alanlarda bulunan verilerin özellikleri birbirine uyumlandırmaya çalışılır. Burada genellikle kaynak ile hedef arasında bulunan istatistiksel farklar analiz edilir ve bu farkları minimize etmek için Maksimum Ortalama Tutarsızlık (Maximum Mean Discrepancy - MMD)[3] ve Korelasyon Uyumlaştırılması (Correlation Alignment - CORAL)[4] gibi mesafe tabanlı yaklaşımlara dayalıdır. İki alana ait dağılımların arasındaki farkı bir kayıp fonksiyonu olarak ifade ederek bir regularizasyon problemi olarak değerlendirilerek matematiksel optimizasyon gerçekleştirilir.

2.3.3. Çekişmeli Alan Uyarlama

Klasik yöntemlere alternatif olarak yapay zeka alanında daha işlem kapasitesi yüksek makineler ile daha farklı model yapıları kurulmaya başlandıktan sonra bu alanda yeni yöntemler denenmeye başlanmıştır. Bunlara örnek vermek gerekirse 2015 yılı sonrası üretken çekişmeli ağların (generative adversarial networks - GAN) yaygınlaşması ile çekişmeli öğrenme algoritmaları alan uyarlamada kullanılmaya başlanmıştır[32]. Bu yöntemde temel olarak alan ayırıştırıcı (domain discriminator) ve özellik çıkarıcı (feature extractor) olmak üzere iki farklı bileşen vardır. Özellik çıkarıcı, ayırıştırıcının tahmin yeteneğini şaşırtmak için farklı özellikler çıkartarak farklı çeşitlerde veriyi temsil edecek bir görev görür. Ayırıştırıcı, özellik çıkarıcının çıktılarını ayırt etmeye çalışır ve bu konuda ikisi birbiri ile yarış haline girerler. Bu yaklaşımın en yaygın bilinen örneği Çekişmeli Alan Sinir Ağıdır (Domain-Adversarial Neural Network – DANN)[33] .

2.3.4. Piksel Düzeyinde Uyarlama

Anlamsal bölütleme alanında her bir piksel değeri için tahminleme yapıldığı için buna yoğun tahminleme denmektedir. Yoğun tahminleme problemlerinde, görüntüye ait özellikler ile yapılan tahminlemelerde kullanılmaktan ziyade bunu gerçekleştirme için piksel tabanlı tahminleme yapabilmek için yeni yöntemler geliştirilmiştir. Bu tarz yöntemlerde, kaynak alana ait görüntüler, hedef alanı benzetilerek sanki onları taklit edecek şekilde değiştirilerek eğitimde kullanılması amaçlanmaktadır. Bunu gerçekleştirmek amacıyla da adına üreteç (generator) denilen modeller oluşturulmaktadır. Bu üreteçler model eğitimi süresince eğitime tabi tutulup sanki kaynak alana ait görüntüyü, hedef alana benzeterek eğitime sokar bu sayede hedef alana benzeyen görüntüler eğitime sokulduğu için anlamsal bölütleme modelleri hedef alana benzer görüntüler ile eğitildiği için burada eğitilen model hedef alanda yapılan tahminlemelerde iyi performans göstermesi amaçlanır. Burada eğitilen üreticinin eğitilmesi amacıyla da bir ayrıştırıcı kullanılmaktadır. Bu ayrıştırıcı tıpkı özellik çıkarıcı-ayrıştırıcı yapısındaki gibi üreticiyi kandırarak onun çıktılarının daha da hedef alanda verilerine benzetmesine sebep olur. Bu sayede üreticinin performansı artar. Bu tarz modellere CycleGAN[34] tabanlı alan çevirmesi (domain translation) modelleri denmektedir. Stil benzetimi hedeflenir ve görsel tutarlılık sağlarlar. Fakat bu yöntemlerde görsel bozulmalar oluşabilmektedir bu da ana eğitilen modelin performansının düşmesine sebep olabilmektedir [31].

2.3.5. Frekans Tabanlı Alan Uyarlama

2.3.5.1. Fourier Alan Uyarlama

Fourier alan uyarlaması (Fourier Domain Adaptation - FDA) tekniği, kaynak veri ile hedef veri arasında frekans düzeyinde bir geçiş sağlayarak iki alan arasındaki alan kaymasını azaltmayı amaçlamaktadır [13]. Bu yöntemde öncelikle hedef alan ve kaynak alan veri kümesindeki görüntülerin Fourier dönüşümleri alınarak işlemlere başlanılmaktadır. Fourier dönüşümü alındıktan sonra genlik ve faz olarak iki farklı çıktı elde edilmektedir. Frekans alanında incelersek faz bilgisi genel olarak görüntülerin kenar geçişleri, objeler gibi mekansal yapıları içermektedir. Bunun aksine genlik spektrumu ise görüntüye ait olan stil, ton farkı, renkler, aydınlatma gibi bilgileri

içermektedir. Bu bilgiler ışığında hedef alan hakkında bilgi sahibi olmuş olabiliriz. Elde edilen bu bilgiler ışığında Fourier alan uyarlaması gerçekleştirilebilir. Kaynak alandaki görüntülere, hedef alan görüntülerinin genlik spektrumdaki merkez alanda bulunan düşük frekanslı bileşenleri bir pencere yardımı ile yer değiştirme yapılmaktadır. Bu sayede kaynak görüntüsü hedef görüntüsü arasında bir stil benzetim işlemi sağlanabilmektedir. Bu işlemleri yapabilmek için her iki görüntünün Fourier dönüşümünü almak için Hızlı Fourier Dönüşümü (Fast Fourier Transform – FFT) uygulamak gerekmektedir[35].

$$\mathcal{F}(x)(m, n) = \sum_h \sum_w x(h, w) \cdot e^{-j2\pi((h/H)m + (w/W)n)} x, \quad j^2 = -1 \quad (2.8)$$

FFT ile Fourier alınarak merkez bölgedeki frekanslar iki alan arasında değiştirilerek yeni genlik spektrumları elde edilmiş olur. Sonrasında ters Fourier dönüşümü ile görüntüler geri elde edilir

$$M_\beta(h, w) = \mathbb{1}_{(h,w) \in [-\beta H: \beta H, -\beta W: \beta W]} \quad (2.9)$$

$\beta \in (0,1)$ hiperparametresi, genlik spektrumun ortasında ne kadarlık bir bölgenin pencere ile kaynak ile hedef arasında transfer edileceğini belirlemektedir.

Eğer ki β hiperparametresi $\beta=0$ olarak seçilirse herhangi bir alan uyarlama yapılamaz çünkü pencerenin boyutu 0 olur. Ancak $\beta=1$ olarak seçilirse hedef alana ait görüntünün tüm genlik spektrum bileşenleri aktarılmış olur bu da mantığa ters olmaktadır. Tipik olarak β hiperparametresi düşük frekansları kapsayacak şekilde bir değer olarak seçilir. Aksi takdirde β yüksek seçilirse stil aktarımı sırasında yüksek genlikler sebebiyle görüntülerde bozulmalar olabilmektedir.

$$x^{s \rightarrow t} = \mathcal{F}^{-1} \left(M_\beta \circ \mathcal{F}^{\mathcal{A}}(x^t) + (1 - M_\beta) \circ \mathcal{F}^{\mathcal{A}}(x^s), \mathcal{F}^{\mathcal{P}}(x^s) \right) \quad (2.10)$$

Hedef stilli kaynak görüntüsünün elde edilim denklemi verilmiştir. Kaynak alan görüntüsünün faz spektrumu korunurken genlik spektrumunda bir pencere ile hedef alan görüntüsünün genlik spektrumu ile değiştirilmektedir.

- x^s : Kaynak alan görüntüsü (s - Source)
- x^t : Hedef alan görüntüsü (t - Target)

- $x^{s \rightarrow t}$: Hedef stilli kaynak alan görüntüsü
- $\mathcal{F}^{\mathcal{A}}(x^s)$: Genlik spektrumu
- $\mathcal{F}^{\mathcal{P}}(x^s)$: Faz spektrumu
- $\mathcal{F}^{-1}$: Ters Fourier dönüşümü

Fourier alan uyarlamasının amacı düşük frekanslı değişimlerin dahi aktarılacak derin öğrenme modellerinin performansını arttırmaktır. Herhangi bir karmaşık model kurgulamaya veya CycleGAN’da olduğu gibi herhangi bir çekişmeli öğrenme metoduna ihtiyaç duymamaktadır [34]. Fourier dönüşümü ile bu işlemler başarılı bir şekilde gerçekleştirilebilmektedir. Derin öğrenme boru hatlarına kolayca entegre edilebilir bir yapıya sahiptir. Yang ve arkadaşlarının yaptığı bu çalışma ile bir çok çekişmeli öğrenme modellerinden daha iyi performans elde edilmiştir [3]. Bu başarılar basit bir matematiksel model ile sağlanması alan uyarlama alanında ilgi çekicidir.

2.3.5.2. Gauss Fourier Alan Uyarlama

Fourier alan uyarlamasında frekans alanındaki değişimler için tasarlanan pencere kare bir pencere iken Gauss Fourier alan uyarlamasında bu pencere bir gauss penceresidir. Temel farkı keskin frekans geçişleri yerine daha yumuşak geçişli bir filtre ile stil aktarımının daha yumuşak bir şekilde yapılmasıdır. Fikir olarak gauss penceresi ile yapılacak olan stil aktarımı daha az gürültü içerecektir. Örneğin: renk geçişleri daha iyi gözükülecektir. Frekans aktarımı sebepli görüntüde dalgalanmalardan kaçınılacaktır. Bu gibi nedenler yüzünden böyle bir gauss penceresi tasarımı Basak ve Yin tarafından 2023 yılında yayınladıkları makalede kullanılmıştır.[36] Bu yöntem sayesinde hedef görüntü stiline daha doğal bir yaklaşım yapılmış olunur.

$$M\sigma(h, w) = e^{\left(-\frac{(h - H/2)^2 + (w - W/2)^2}{2\sigma^2}\right)} \quad (2.11)$$

Denklem 2.11 de verilen Gauss FDA yönteminde kullanılan Gauss penceresinin parametresi olan standart sapma (σ) tasarım için önem arz etmektedir. σ parametresi ile oluşturulacak olan pencerenin düzlem üzerinde yayılımı kontrol edilmektedir. (σ) değeri küçük seçilirse daha dar bir pencere, daha geniş seçilirse daha geniş bir pencere elde edilir.

Burada önemli olan en temel nokta bu tarz bir pencerenin türevlenebilir olmasıdır. Bu sayede bu pencere öğrenilebilir bir maske olarak eğitime girdi yapılabilir. Eğitim süresince Gauss penceresi parametreleri öğrenilerek veri kümesi için en optimal maskeye yaklaşılmaya çalışılmaktadır. Liu ve arkadaşlarının 2021 [4] yılında kolonoskopi görüntüleri ile yaptıkları çalışmada farklı cihazlara ait görüntülerin arasında renk ve aydınlatma problemini çözmek için Gauss penceresi ile Fourier alan uyarlaması uygulanmıştır. Gauss penceresi ile yapılan daha yumuşak geçişli spektrum transferi kullanımı ilerleyen yıllarda da problemleri çözmek için kullanılmıştır[37].

2.3.5.3. Frekans Tabanlı Optimal Stil Karışımı

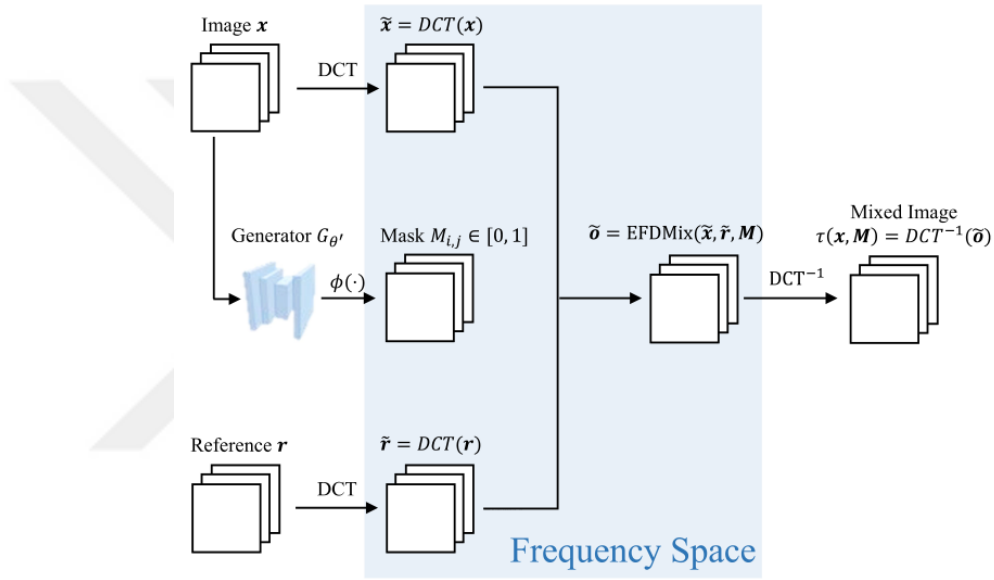
Frekans Tabanlı Optimal Stil Karışımı aslında bir alan genellemesi (domain generalization) problemleri için geliştirilmiş olan bir yöntemdir [8]. Bu yöntem ile bir veya daha çok kaynak alandan öğrenim görmüş olan modelin hiç görmediği bir alanda iyi bir performans göstermesi amaçlanmaktadır. Bunun için veriler üzerinde stil karıştırması (style mixing) uygulanmaktadır. Alan genelleştirmede hedef alana ait veriler eğitimde kullanılmamaktadır. Bu sebeple aslında denetimsiz alan uyarlama yöntemlerine benzemektedir. Modelin eğitilirken hedef alan ait veri ile eğitilmediğinden dolayı iyi performans sergileyebilmesi amacıyla alanlardan bağımsız (domain-invariant) özelliklere sahip olması gerekmektedir. Bu sebeple fazlaca farklı stil çeşitleri ile modellerin eğitilmesi gerekmektedir. FOSMix yöntemi ile bu stil karıştırma veya çeşitlendirme işlemleri frekans alanında yapılmaktadır. FOSMix yöntemiyle, her bir görüntüye ait içerik korunarak görüntüye ait ton, aydınlatma gibi stil farklılıkları rastgele veya optimal bir şekilde çoklanarak nispeten farklı ortamlardan veriler varmış gibi eğitim senaryosu geliştirmek amaçlanmaktadır. Model eğitime getirilen bu yaklaşım sayesinde alan bağımsızlaştırma sağlanmış olur. FOSMix yöntemi üç temel aşamadan oluşur. Bunlar tam karışım, optimal karışım ve tutarlılık düzenlemesidir.

İlk aşama tam karışım aşamasıdır. Hedef alana ait veri kümesindeki görüntülerin stil karıştırması yapılırken bu işlemlerin maksimum düzeyde yapıldığı kısımdır.

Optimal karışım ise Tam karışım ile yapılan karışımdan ziyade optimal bir maske ile bu stil karıştırma işleminin yapılması kısmıdır. Tam karıştırma ile tüm frekanslar karıştırıldığı için görüntülerde bozulmalar gözükabilmektedir. Bundan kaçınmak

amacıyla anlamsal bölütleme için gereken kısımlar korunurken geriye kalan frekans bileşenleri karışıma girdi yapılır. Modelin ihtiyaç duyduğu yapısal bilgiyi kaybetmeden veri kümesi stil olarak çeşitlendirilir.

FOSMix yöntemi model eğitiminde kullanılırken modelin orijinal olarak eğitildiği kaynak alan görüntüsü ile stil karışımı yapılan görüntü karşılaştırılır. Bu karşılaştırma sonucu elde edilen tutarlılık kayıp fonksiyonu ile modele farklı tür stillerde girdi yapılırsa dahil anlamsal olarak aynı sonucu tahminlenmesi amaçlanmaktadır. Bu kayıp fonksiyonu ile eğitim modelinin kararlaştırılması hedeflenmektedir.



Şekil 2.10: FOSMix alan uyarlama akışı şeması [5].

Şekil 2.10 incelendiğinde girdi görüntüsünün hem FM hem de OM kısmına girdi yapıldığı gözlemlenmektedir. Kaynak alana ait görüntü Ayırık Kosinüs Dönüşümü (Discrete Cosine Transform) ile frekans alanına aktarımı gerçekleştirilmektedir[38].

$$\tilde{x} = DCT(x), \tilde{r} = DCT(r) \quad (2.12)$$

x : Girdi Görüntüsü

r : Referans Görüntü

$$M = G_{\theta'}(\cdot) \quad (2.13)$$

Denklem 2.13'te verilen M bir optimal maskedir. Bu optimal maske genellikle bir evrişimsel sinir ağı tarafından oluşturulmaktadır. Optimal olarak tahmin edilen maske her bir frekans değeri için stil karışım oranını belirlemektedir. Maskenin 0 olan kısımları yalnızca kaynak alandan, 1 olan kısımları yalnızca hedef alandan, ara değerler ise iki alanın karışımı olmak üzere açıklanabilir.

$$\tilde{o} = EFDMix(\tilde{x}, \tilde{r}, M) \quad (2.14)$$

EFDMix yapısı, girdi yapılan kaynak alan görüntüsü ile hedef alan görüntüsünü bir maske yardımıyla karıştırmayı sağlamaktadır.

$$\tilde{o}_{i,j} = (1 - M_{i,j}) \cdot \tilde{x}_{i,j} + M_{i,j} \cdot \tilde{r}_{i,j} \quad (2.15)$$

Denklem 2.15'e bakıldığında karışım denklemi verilmiştir. Bu karışım denklemi ile bir stil transferi yapmak mümkün hale gelmektedir.

$$\tau(x, M) = DCT^{-1}(\tilde{o}) \quad (2.16)$$

Stil karışımı yapıldıktan sonra tekrar geri ayrı kosinüs dönüşümü yapılarak hedef alan stilini taşıyan karışık görüntü $\tau(x, M)$ elde edilmiş olur.

2.4. Veri Kümeleri

Literatürde otonom araçlarda anlamsal bölütleme probleminin başarımlarını incelemek amacıyla çeşitli iyi veya seyrek etiketlenmiş çok sayıda veri kümesi bulunmaktadır. Yaygın olarak bu alanda kullanılan veri kümeleri: Cityscapes[2], SYNTHIA[39], Pascal VOC[40], KITTI[41], GTA5[1], ADE20K[42], BDD100K[43] ve CamVid[44] gibi çok çeşitli veri kümeleri yer almaktadır. Bu bahsedilen veri kümeleri, farklı coğrafyalarda, alanlarda, zamanlarda, hava koşullarında ve ayrıca içerdiği etiketler bakımından birbirinden farklılaşmaktadırlar.

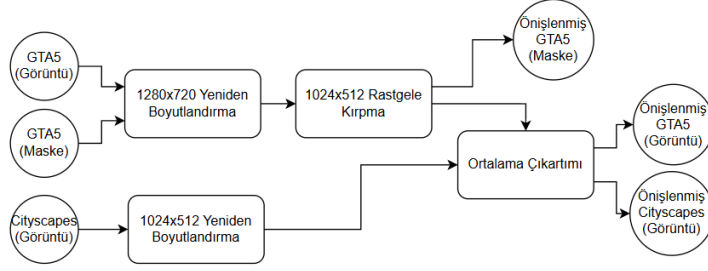
3. YÖNTEM

Bu tez kapsamında FDA[3], GFDA[4] ve FOSMix[5] olmak üzere üç farklı gözetimsiz alan uyarlama yöntemleri denenmiştir. Denenen yöntemlerde parametreler sabit kalsın diye sadece alan uyarlama yöntemi değiştirilerek denemeler yapılmıştır. Veri ön işleme ve model eğitim adımları her bir yöntem için farklılık göstermemektedir.

3.1. Veri Ön İşlemesi

Kaynak alana ait görüntülerini içeren veri kümesi GTA5[1] olarak seçilmiştir. Hedef alana ait veri kümesi ise Cityscapes[2] olarak seçilmiştir. İki farklı veri kümesi girdi olarak alınmıştır. GTA5 veri kümesi 24,966 tane sentetik görüntü içermektedir ve boyutları orijinalde 1914x1052'tir. Bu çözünürlük değeri eğitim süresince düşürülüp kırılmaktadır. Kaynak alan görüntüsü çözünürlüğü 1280x720 olacak bikübik(bicubic) interpolasyon yöntemi ile tekrardan boyutlandırılır. Ground-truth kaynak alan maskesi ise en yakın komşu interpolasyonu ile aynı boyuta düşürülür. Hem kaynak alan görüntüsü hem de kaynak alan maskesi rastgele bir şekilde kesilerek boyutu 1024x512'ye düşürülür. Orijinal GTA kümesi 33 farklı sınıf etiketlerine sahiptir. 19 tanesi Cityscapes veri kümesi ile uyduğundan sadece o sınıflar dikkate alınarak diğerleri devre dışı bırakılır. En son olarak kaynak alan görüntüsünden ortalama ImageNet[23] yarışmasında kullanılan veri kümesine ait olan ortalama değerler çıkartılmaktadır.

Cityscapes verisi ise orijinal boyutu 2048x1024'ten, 1024x512'ye bicubic interpolasyon yöntemi ile yeniden boyutlandırılır. Fakat Cityscapes veri kümesi için herhangi rastgele bir kesme uygulanmamaktadır. GTA5 veri kümesinde olduğu gibi ImageNet veri kümesine ait ortalama değerler çıkartılmaktadır.



Şekil 3.1: Veri ön işleme Akış diyagramı.

3.2. Alan Uyarlama

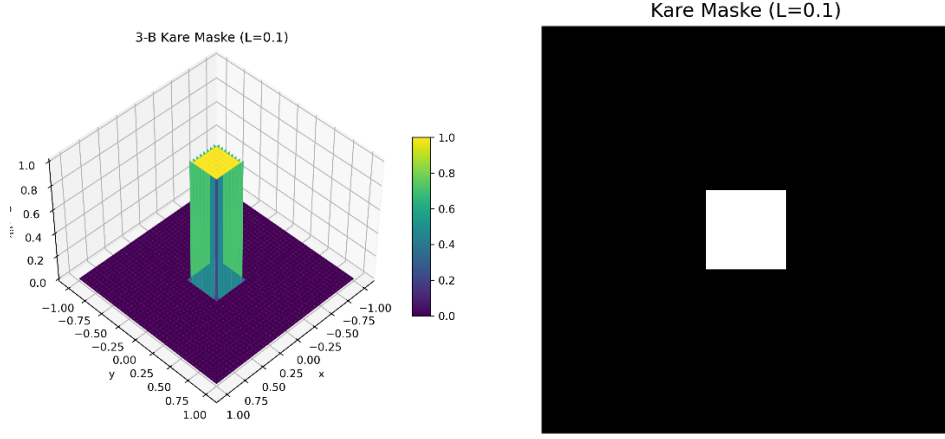
Önceden de bahsedildiği üzere 3 farklı alan uyarlama yöntemi ile kaynak ile hedef arasında bir stil aktarımı yapılmaktadır.

3.2.1. Fourier Alan Uyarlaması

Kaynak alan x_s görüntüsü ile hedef alan x_t görüntülerine rFFT uygulanmaktadır[45]. rFFT sayesinde frekans alanında spektrum yarıya düşmektedir. Hem bellek hem de hesap maliyeti bu sayede yarıya inmektedir. Genlik ve faz spektrumu olarak ikiye ayrılır. β değeri maksimum değeri 1 olacak şekilde sınırlandırılır. Sonrasında beta değerine göre genlik spektrumu üzerinde oluşturulan pencere ile onun merkezlendiği frekans değerleri elde edilir. Sonrasında frekans transferi işlemi gerçekleştirilir. Güncellenmiş genlik $\widetilde{A}_s$ ile eski faz değeri ϕ_s birleştirilerek hedef alan stiline sahip kaynak alan görüntüsü genlik spektrumu oluşturulur[3]. Sonrasında iFFT ile ters Fourier dönüşümü alınarak hedef alan stiline kaynak alan görüntüsü elde edilmiş olunur.

$$M_{\beta}(k_x, k_y) = \begin{cases} 1, & \text{eğer } |k_x| \leq \beta \text{ ve } |k_y| \leq \beta \\ 0, & \text{diğer} \end{cases} \quad (3.1)$$

$$\widetilde{A}_s = \begin{cases} A_t(k), & \text{eğer } k \in \text{düşük frekanslar (maskeli alan)} \\ A_s(k), & \text{diğer} \end{cases} \quad (3.2)$$



Şekil 3.2: Kare maske.

Şekil 3.2’de genlik spektrumunda bulunan düşük frekansları değiştirmek için kullanılan kare pencere örneği verilmiştir.

3.2.2. Gauss Fourier Alan Uyarlaması

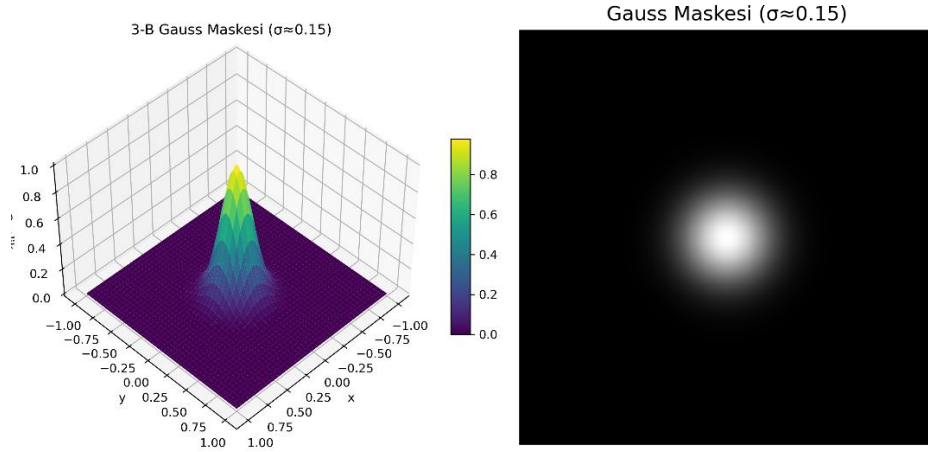
Gauss Fourier Alan Uyarlamasında frekans spektrumunda kare bir pencere yerine yumuşak geçişli bir pencere kullanılmıştır[4]. Bu pencere türevi alınabilir bir şekilde tasarlanmıştır. Türevi alınıyorsa aynı zamanda model eğitim sürecinde de bu pencere öğrenilebilir olur. Pencere tasarımına ait detaylar Denklem 3.1, 3.2, 3.3 ve 3.4’te verilmiştir. k_x x-ekseninde, k_y y-ekseninde bir katsayıyı temsil ederler. Denklem 3.2 bir sigmoid fonksiyonudur.

$$k_x \in \left[-\frac{W_f}{2}, \frac{W_f}{2} - 1 \right] \quad (3.3)$$

$$k_y \in \left[-\frac{H}{2}, \frac{H}{2} - 1 \right] \quad (3.4)$$

$$\sigma(z) = \frac{1}{1+e^{-z}} \quad (3.5)$$

$$M_{\{\beta, \sigma\}(k_y, k_x)} = \sigma(10(\beta - |k_y|)) \sigma(10(\beta - |k_x|)) + \left(1 - \sigma(10(\beta - |k_y|)) \sigma(10(\beta - |k_x|)) \right) * \exp\left(-\frac{k_y^2 + k_x^2}{2\sigma^2 + 10^{-6}}\right) \quad (3.6)$$



Şekil 3.3: Gauss maskesi.

Şekil 3.3'te genlik spektrumdaki düşük sinyalleri kare maskeye göre daha yumuşak geçişler ile transfer etmek için oluşturulan Gauss maske verilmiştir.

3.2.3. Frekans Tabanlı Optimal Stil Karışımı

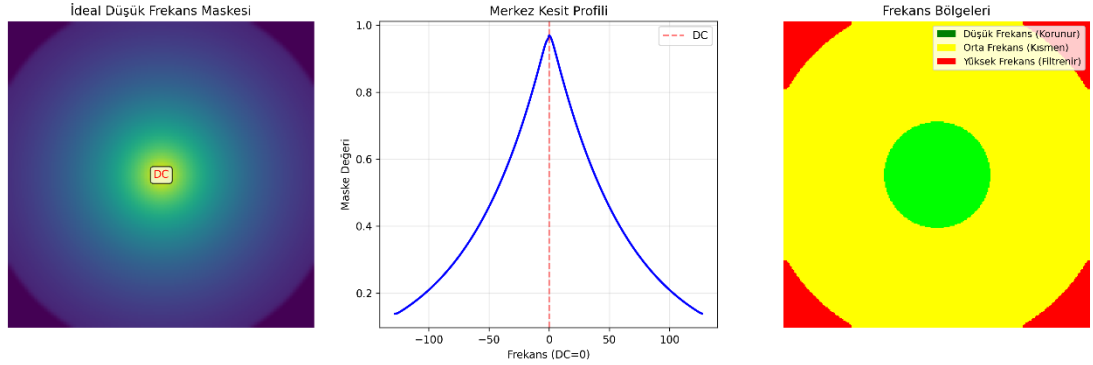
FOSMix yöntemi ile kaynak alan görüntüleri ile hedef alan görüntülerinin stilleri karıştırılarak birbirine benzetim işlemi yapılmaktadır. Frekans alanında iki seviyeli karışım metodu tanımlar[5]. Bunlar tam karışım (Full Mix – FM) ve optimal karışım (Optimal Mix - OM)'dir. Bu tez kapsamında FOSMix yöntemi ile karışım yaparken düşük frekanslı bölgelere tıpkı Fourier alan uyarlamasında yapıldığı gibi kare ve gauss pencere tasarımları yapılmıştır. Ayrıca optimal bir karışım sağlamak amacıyla hem FOSMix çalışmasında kullanılan optimal maskeyle hem de U-Net tabanlı bir maske ile çalışılmıştır. Bunlara ek olarak orijinal makalede kullanılan veri ön işleme adımları birebir uygulanmıştır. Yani tam karışım ve optimal karışım aynı anda uygulanmıştır ve burada yapılan optimal karışım, orijinal makalede kullanılan maske ve UNet maskesi ile farklı eğitimler gerçekleştirilmiştir. Denklem 3.7 ve 3.8'de sırasıyla tam karışım ile optimal karışıma yer verilmiştir.

$$A_{FM} = A_t \quad (3.7)$$

$$A_{OM} = (1 - M_\theta)A_s + M_\theta A_t \quad (3.8)$$

M_θ Maskesi, kare, gauss ve optimal maske olarak seçilerek eğitimler gerçekleştirilmiştir.

Yapılan bu çalışmada hem DCT ile hem de FFT ile çalışma yapılmıştır. FFT ile gerçekleştirilen eğitimlerde sadece maske ile sınırlı alan ile karışım yapılmıştır. DCT ile ise orijinal makalede yapıldığı gibi öğrenilebilir bir maske sayesinde optimal karışım ve tam karışım ile birlikte model eğitimi gerçekleştirilmiştir.



Şekil 3.4: FOSMix optimal maske.

Şekil 3.4’te FOSMix yönteminde stil transferi işlevini yerine getirmek için öğrenilebilir düşük frekans maske örneği verilmiştir.

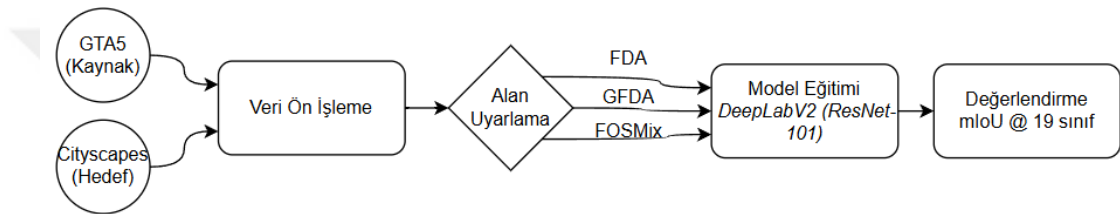
Geliştirilen FOSMix yönteminde öğrenilebilir maske U-Net-HalfMask’tır[28]. Evrişim, havuzlama, yukarı örnekleme kademeleri içerir. rFFT’nin doğası gereği yarım bir maske önerisi yapılmıştır. Kare ve Gauss penceresine göre daha esnek bir maske yapısı oluşturulmaya çalışılmıştır. Ayrıca yapılan bu çalışmada, FOSMix yönteminde kullanılan orijinal optimal maskeye de yer verilmiştir. Bu optimal maske ise UNet[28]’ e kıyasla daha az katmanlı ve daha az nörona sahip bir konvolüsyonel ağıdır. Gauss maskesi kullanılan FOSMix kısaca GFOSMix, kare maske kullanan FOSMix ise SFOSMix olarak isimlendirilmiştir.

3.3. Model Eğitimi

Ön işlemeden geçmiş kaynak alan ve hedef alan veri kümeleri tanımlanır. Tanımlanan bu veri kümeleri paket paket model eğitim döngüsüne girdi olarak verilir. Görüntü çözünürlükleri yüksek olduğundan hafızada Rahat bir performans olması amacıyla batch sayısı 1 olarak seçilmiştir. Eğitim için toplam adım sayısı 150.000 olarak

belirlenmiştir. Her bir iterasyonda batch=1 olduğu için bir çift veri NVIDIA RTX 4060Ti GPU 16GB platformuna taşınmaktadır.

Eğitim döngüsünde her iki alana ait veri kümelerinden gelen görüntüler istenen alan uyarlama tekniğine sokularak bir stil transferi gerçekleştirilir. Bu alan uyarlama çıktısı olarak elde edilen hedef görüntü stilini barındıran kaynak görüntü modele girdi yapılmak üzere tanımlanmaktadır. Veri kümesine ait ortalamalar kullanılarak tanımlanan hedef görüntü stilini içeren kaynak görüntüden bu ortalama değeri çıkartılmaktadır. Veri model girdisi yapılmak için hazır duruma getirilmiştir. Örnek model akış şemasına Şekil 3.5’te yer verilmiştir.



Şekil 3.5: Model akış şeması.

3.3.1. Model Parametreleri

Anlamsal bölütleme işini güncel olarak iyi başaran modellerden biri olan DeepLab mimarisinin version 2 modeli seçilmiştir[9]. Backbone olarak ResNet-101 seçilerek eğitimler bu mimari ile yapılmıştır[26]. Model’e ait hiperparametreler, modeli güncellemek için kullanılan kayıp fonksiyonları ve eniyileciler ile ilgili bölümlerde bahsedilmiştir.

3.3.2. Öğrenme Oranı

Model eğitim sürecinde ağırlıkların ne kadar güncellenmesi gerektiğini belirten öğrenme oranı hiperparametresi tüm anlamsal bölütleme ağlarında başlangıç değeri olarak $lr = 2.5 \times 10^{-4}$ olarak seçilmiştir.

Öğrenme oranı parametresi adım sayısına bağlı olarak güncellenmiştir

$$lr_i = lr_0 \left(1 - \frac{i}{n_i}\right)^{kat} \quad (3.9)$$

i: Adım sayısı

ni: Toplam adım sayısı

Denklem 3.9'da $\text{kat}=0,9$ olarak seçilmiştir.

3.3.3. Eniyileyiciler

Anlamsal bölütlemeye kullanılacak model için ayarlanan eniyileciyi (optimizer) Stokastik Gradyan İnişi (Stochastic Gradient Descent - SGD) algoritmasıdır[46].

SGD için momentum=0,9 olarak seçilmiştir. L2 ağırlık ceza puanı ise 5×10^{-4} seçilmiştir.

Gauss Fourier Alan Uyarlamasında, Gauss maskesi için eniyileyici Adam olarak seçilmiştir. Adam eniyileyicisi Gauss maskesine ait β ve σ parametrelerini optimize etmeye çalışır[47]. Bu eniyileyici için başlangıç öğrenme oranı $lr_{(\beta,\sigma)} = 10^{-3}$ seçilmiştir. Her 25000 adımda bu öğrenme oranı 0,1 oranında düşürülmektedir. Bu yapılan işleme StepLR denir.

$$lr_{(\beta,\sigma)} = lr_{(\beta,\sigma)} \times 0.1 \quad (3.10)$$

3.3.4. Eğitim Adımları ve Batch Büyüklüğü

Eğitim döngüsünde batch büyüklüğü 1 örnek içeren mini batch kullanılmıştır. Toplam eğitim için ayrılan adım sayısı 150.000 olarak belirlenip 100.000 adıma ulaşıldığında modelde herhangi bir iyileşme gözükmemesi durumunda erken durdurma özelliği aktif edilmiştir.

3.3.5. Model Kayıt Süreci

Eğitim döngüsünün her 2500 adımında bir model ağırlıkları kaydedilerek istenmeyecek güç, hafıza problemlerine karşı önlem alınmaktadır.

3.3.6. Kayıp Fonksiyonları

FDA[3] ve Gauss FDA[4] Alan uyarlamaları yöntemlerinde kullanılan kayıp fonksiyonu iki farklı fonksiyondan oluşur. Bunlar çapraz entropi (Denklem 3.11) ve hedef alan için entropi kaybıdır (Denklem 3.12).

$$L_{seg}^{kaynak} = -\sum_{c=1}^C y_c \log\left(\frac{e^{f_c}}{\sum_{j=1}^C e^{f_j}}\right) \quad (3.11)$$

$$L_{ent}^{hedef} = -\sum_{c=1}^C p_c \log(p_c), \quad p_c = \frac{e^{f_c}}{\sum_{j=1}^C e^{f_j}} \quad (3.12)$$

p_c : softmax[48] fonksiyonu ile c sınıfı için olasılık belirtmektedir. Logaritması alınınca log-softmax elde edilmiş olur. Entropi değeri 19 sınıf için normalize edilerek kullanılmaktadır.

$$\widetilde{L_{ent}^{hedef}} = \frac{L_{ent}^{hedef}}{\log(C)} = \frac{L_{ent}^{hedef}}{2.9444} \quad (3.13)$$

Entropi fonksiyonunun türevlenebilir olması amacıyla karesi alınır ve ufak bir sabit ile toplanarak 0 değerlerinden kaçınılmaktadır.

$$L_{ent} = (\widetilde{L_{ent}^{hedef}}^2 + \epsilon)^n \quad (3.14)$$

Denklem 3.12’de kullanılan n parametresi kötü tahminleri cezalandırmak için kullanılmaktadır.

FDA ve Gauss FDA için toplam kayıp fonksiyonu Denklem 3.13’te verilmiştir. Denklem 3.13’te kullanılan $\lambda_{ent}=0.005$ parametresi ile entropi kaybının etkisi kontrol edilebilmektedir.

$$L_{toplaml} = L_{seg}^{kaynak} + \lambda_{ent} L_{ent}^{hedef} \quad (3.15)$$

Entropi sayesinde modelin daha kesin tahminler yapması amaçlanmaktadır.

FOSMix yönteminde yine aynı kayıp fonksiyonlarına yer verilmektedir fakat FM ve OM yöntemlerini kullandığımızda bunlardan elde edilen consistency_loss adında bir kayıp fonksiyonu daha kullanılmaktadır.

$$\hat{y} = softmax(pred) \quad (3.16)$$

$$L_{spatial}^{OM} = \frac{1}{NHW} \sum_{c,h,w} |\hat{y}_{raw}^{c,h,w} - \hat{y}_{om}^{c,h,w}| \quad (3.17)$$

$$L_{spatial}^{FM} = \frac{1}{NHW} \sum_{c,h,w} |\hat{y}_{raw}^{c,h,w} - \hat{y}_{fm}^{c,h,w}| \quad (3.18)$$

$$L_{spatial} = \frac{L_{spatial}^{OM} - L_{spatial}^{FM}}{2} \quad (3.19)$$

$$L_{DCT}^{OM} = \frac{1}{NHW} \sum_{c,h,w} |D(\hat{y}_{raw}) - D(\hat{y}_{fm})| \quad (3.20)$$

$$L_{DCT}^{FM} = \frac{1}{NHW} \sum_{c,h,w} |\hat{y}_{raw} - \hat{y}_{fm}| \quad (3.21)$$

$$L_{DCT} = \frac{L_{DCT}^{OM} - L_{DCT}^{FM}}{2} \quad (3.22)$$

$$L_{total} = 0.7 * L_{spatial} + 0.3 * L_{DCT} \quad (3.23)$$

Denklem 3.23'te verilen kayıp fonksiyonu modelin alan uyarlama sırasında tahminlerinin hem görüntü düzeyinde hem de frekans düzeyinde stabil ve güvenilir kalmasını sağlar. Yani, alan kaymasına rağmen tutarlı bir davranışa yönelmektedir.

3.4. Model Değerlendirmesi ve Başarım Metrikleri

Model eğitimi bittiğinde kaydedilen model ile hedef alana ait test verileri üzerinde model çalıştırılır. Bu bölümde model hem nitel olarak hem de nicel olarak analiz çıkarımı yapılmaktadır

3.4.1. Test İşlemi

Hedef alana ait görüntüler veri yükleyici ile hafızaya alınır. Hafızadan modele tek tek girdi yapılarak model tahminlemeleri yapılır. Modelden elde edilen tahminler softmax fonksiyonundan geçirilerek 19 sınıftan olasılıksal değeri yüksek olan sınıflar çıkışta elde edilecek hedef alan tahmin maskesinde yerini alır. Bu elde edilen maskeler hem gri tonlamalı hem de renkli bir biçimde kaydedilmektedir.

Eğitim sürecinde olduğu gibi test sürecinde de veri ön işleme olarak ortalama renk vektörü çıkarımı yapılmıştır. Böyle bir işlem ile eğitilen veri ile test edilen verilerin dağılımsal olarak farklılıkları azaltılmaktadır.

3.4.2. Başarım Metrikleri

Modele girdi yapılan hedef alan test görüntülerinden tahminlenen maskeler, hedef alana ait ground-truth maskeler ile karşılaştırmaya sokulur. Bu karşılaştırmada kapsamlı bir değerlendirme yapabilmek için sekiz farklı performans metriği hesaplanmıştır: Intersection over Union (IoU), Piksel Doğruluğu, Ortalama Sınıf Doğruluğu, Frekans Ağırlıklı IoU, Dice Katsayısı, Keskinlik (Precision), Duyarlılık (Recall) ve F1 Skoru. Yapılan değerlendirmede 19 sınıf teker teker bu işlemlere sokulmuş ve her metrik için hem sınıf bazlı hem de ortalama değerler elde edilmiştir. Bu metrikler sayesinde modelin sadece genel başarımı değil, aynı zamanda sınıf bazlı performansı, dengesiz veri dağılımındaki davranışı ve tahmin güvenilirliği gibi farklı açılardan kapsamlı bir değerlendirme yapılmıştır. Özellikle düşük temsil edilen sınıflarda (tren, motosiklet, bisiklet gibi) ve geniş alan kaplayan sınıflarda (yol, bina, gökyüzü gibi) modelin performansı detaylı olarak analiz edilmiştir.

4. TEST VE ANALİZ

Tezin bu bölümünde, Yöntem kısmında anlatımı yapılan model eğitim için kullanılacak olan veri kümelerine. Eğitim sonrası elde edilen tahminleme sonuçlarına, alan uyarlama yöntemlerinin başarımlarının karşılaştırılmasına, öğrenilebilir maske elde edinimine ve bu maskeler ile sabit maskelerin kıyaslamasına, işlem ve performans maliyetlerine değinilecektir

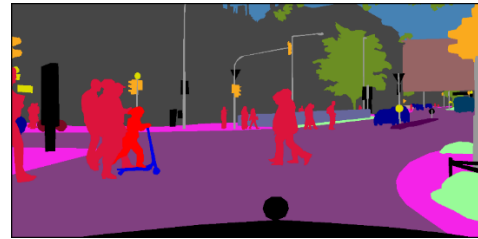
4.1. Veri Kümeleri

4.1.1. Cityscapes

Cityscapes veri kümesi Almanya'nın şehirlerinin yollarında kaydedilmiş yüksek çözünürlüklü gerçek dünya görüntülerden oluşan bir veri kümesidir [2]. 50 farklı şehri kapsayan bu veri kümesinin bir kısmı piksel düzeyinde anlamsal bölütleme uygulamaları için etiketlenmiştir. Bu veri kümesi içermektedir. Cityscapes iyi etiketlenmiş (fine-annotated) veri sayısı toplamda 5000'dir. Görüntülerin çözünürlükleri 2048x1024 olmak üzere 3 kanallıdır.



Şekil 4.1: Cityscapes örnek görüntü.



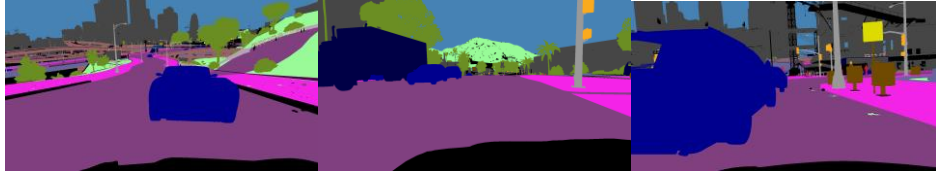
Şekil 4.2: Cityscapes örnek ground-truth.

4.1.2. GTA5

GTA5 veri kümesi, GTA5 video oyunundan toplanan veriler ile hazırlanan sentetik bir veri kümesidir[1]. Sentetik veri olduğundan dolayı bu veri kümesi üretilirken çok çeşitli farklı ortamlarda, farklı mevsimlerde, farklı zamanlarda veriler kaydedilmiştir. Bu sebeple Cityscapes verisi gibi sınırlı değil çok sayıda veri içermektedir. Böyle sentetik verileri üretmek düşük maliyetli ve büyük ölçeklidir. Bu tarz sentetik veriler kolay bir şekilde anlamsal bölütleme için veri azlığını kompanse etmek için temel hazırlamaktadır. GTA5 veri kümesinde toplamda 24966 tane görüntü bulunmaktadır. Bu görüntüler iyi etiketlenmiştir. Etiketlenen 33 sınıftan 19 tanesi Cityscapes verisine ait sınıflar ile uyushmaktadır[2]. Bu sebeple, alan uyarlama için iyi birer çift veri kümesi pozisyonundadırlar. GTA5 veri kümesine ait örnek görüntüler Şekil 4.3 ve 4.4’te incelenebilir.



Şekil 4.3: GTA5 örnek görüntüler



Şekil 4.4: GTA5 örnek ground-truth

Bu iki veri kümesine ait yüzdesel sınıf dağılımları Tablo 4.1’de verilmiştir.

Tablo 4.1: Yüzdesel sınıf dağılımları.

Sınıf	GTA5 (%)	Cityscapes (%)	Fark (%)
Yol	24.5	52.9	+28.4
Bina	20.0	38.0	+18.0
Kaldırım	6.0	9.2	+3.2

Tablo 4.1: Devam.

Yeşil Alan	10.0	17.3	+7.3
Gökyüzü	5.0	5.0	0.0
İnsan	1.0	1.5	+0.5
Direk	0.8	1.0	+0.2
Duvar	2.0	0.5	-1.5
Çit	1.5	0.3	-1.2
Araba	6.0	6.7	+0.7
Diğer Sınıflar	...	...	...

4.2. Performans Kriterleri

Hedef alan test veri kümesi Cityscapes[2] ile model tahminlemeleri sonucu mIoU değerleri farklı frekans tabanlı alan uyarlama yöntemleri ile karşılaştırılmıştır. mIoU metriğinin hesaplanmasının sebebi sınıf bazında tahmin yeteneğini ölçmesidir. 19 sınıfın her biri için IoU hesabı yapılmaktadır. Bu veri kümelerinde veriler eşit dağılmamıştır. Özellikle küçük sınırlı objelere ait etiketli veriler geniş sınırlara sahip olan verilere göre tahminleme sonuçlarında farklılıklar gözlemlenmektedir. mIoU değerine tek olarak bakmak yanıltıcı olabilir onun için bir tablo ile sınıf IoU değerlerini karşılaştırmak daha sağlıklı bir yaklaşım olabilmektedir. Ayrıca geliştirilen alan uyarlama yöntemleri için sırasıyla, piksel doğruluğu, ortalama sınıf doğruluğu, frekans ağırlıklı IoU, ortalama Dice, ortalama keskinlik, ortalama duyarlılık ve ortalama F1 skoru hesaplanmıştır.

4.3. Test Sonuçları

Denetimsiz öğrenme için geliştirilen anlamsal bölütleme modellerinin tahmini Cityscapes[2] hedef alan veri kümesinden hiçbir şekilde model hedef alana ait maske görüntüsü görmeden sadece test görüntüleri ile tahminlemeleri 3.4.1. başlıkta bahsedildiği gibi yapılmıştır. Cityscapes'e ait ground-truth görüntüleri ile yapılan tahminlemeler karşılaştırılıp mIoU skorları elde edilmiştir. Skor hesaplamaları 3.4.2.

başlıkta bahsedildiği gibi yapılmıştır. Tablo 2’de geliştirilen yöntemler kalın harflerle yazılmıştır.

Tablo 4.2: IoU metrik sonuçları.

Sınıf	FD A	GFD A	GFOSmi x	OM FOSMi x	SFOSMi x	FM+O M DCT FOSMi x	UNet FM+O M DCT FOSMi x
yol	86.8	86.5	88.0	83.6	85.3	75.8	74.38
kaldırım	35.1	22.4	26.3	28.1	30.5	21.8	19.40
bina	78.6	76.4	77.3	71.7	79.1	62.1	55.80
duvar	18.5	20.0	13.7	15.9	16.2	8.0	11.67
çit	25.9	26.3	22.0	9.6	24.1	13.4	7.8
direk	28.9	29.1	27.7	22.0	32.0	18.0	18.6
lamba	32.0	30.5	34.1	21.9	29.2	14.7	20.0
tabela	23.1	19.6	19.8	19.2	21.1	7.0	11.5
yeşil alan	82.3	82.4	82.4	71.8	83.2	62.6	57.0
arazi	36.4	35.5	35.9	19.7	34.2	11.4	20.9
gökyüzü	74.0	71.0	73.6	77.5	72.9	73.6	65.2
insan	58.8	58.1	60.7	47.8	57.1	46.3	38.1
sürücü	29.7	31.2	27.8	21.1	27.4	20.0	18.8
araba	72.9	79.6	73.7	59.3	73.3	60.2	65.6
kamyon	21.5	33.4	23.4	16.4	28.4	27.0	7.6
otobüs	35.8	36.7	32.9	21.5	38.2	38.1	25.0
tren	1.8	1.8	1.5	1.0	1.5	3.0	0.6
motor	22.4	30.7	28.9	11.3	22.3	19.5	20.4
bisiklet	39.6	27.5	39.6	28.4	38.4	8.23	3.6
mIoU (%)	42.3	42.03	41.53	34.10	41.81	31.10	28.5

Tablo 2 incelendiğinde, FDA[3] yönteminin birçok sınıfta geliştirilen diğer yaklaşımlara kıyasla daha iyi performans gösterdiği ve önemli ölçüde üstünlük sağlamaya devam ettiği görülmektedir. FDA yönteminin zayıf kaldığı sınıflar, genellikle sınırları geniş olanlardan ziyade dar alanlara sahip sınıflardır. Örneğin; kamyon, motosiklet, otobüs ve sürücü sınıflarında diğer yöntemlerin daha yüksek başarı gösterdiği gözlemlenmiştir. Buna karşın, yollar, binalar, yeşil alanlar ve çevre gibi geniş sınırlara sahip sınıflarda FDA oldukça yüksek performans sergilemektedir.

Gauss penceresi ile yapılan FDA uyarlaması (Gauss FDA), Tablo 2'nin üçüncü sütununda incelendiğinde, bazı sınıflarda anlamlı ölçüde performans düşüşleri gözlemlenmiştir. Bu sınıflar arasında kaldırım, gökyüzü ve bina gibi geniş bölütleme alanlarına sahip olanlar bulunmaktadır. Ancak Gauss FDA'nın tren, kamyon ve motosiklet gibi araç sınıflarında katkı sağladığı ve daha başarılı sonuçlar ürettiği dikkat çekmektedir.

Sabit kare bir maske ile gerçekleştirilen FOSMix[5] stil transferi yöntemi incelendiğinde, FDA'ya kıyasla bazı sınıflarda iyileşmeler sağlandığı görülmektedir. Bu iyileşme özellikle bitki örtüsü ve yolda engel olarak bulunan direkler gibi sınıflarda belirgindir. Sabit bir maske yerine daha yumuşak geçişler sağlayan Gauss penceresi ile gerçekleştirilen FOSMix uyarlaması, FDA ile benzer sonuçlar üretmiş ve insan ile arazi sınıflarında olumlu katkılar sağlamıştır.

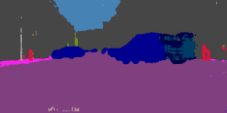
U-Net tabanlı yarı-optimal maske ile eğitilen FOSMix modeli, sınıflar bazında düşük performans göstermiştir. Bu düşüklüğün sebebi olarak, eğitim adım sayısının sabit 100.000 olarak belirlenmesi sonucu maskenin aşırı öğrenme veya yetersiz öğrenme durumlarıyla karşı karşıya kalması gösterilmektedir. Bu durum, öğrenilebilir maskelerin eğitim adım sayısına ve süresine duyarlı olduğunu ortaya koymaktadır. Bu nedenle, daha iyi bir optimizasyon stratejisi ile bu tarz maskelerin, sabit maskelere kıyasla daha etkili düşük frekans seçimleri gerçekleştirebileceği değerlendirilmektedir.

FDA yöntemi, genel anlamda yüksek örnek sayısına sahip sınıflarda üstün performans sergilerken, düşük örnek sayısına sahip sınıflarda diğer yöntemler tarafından geçilmiş görünmektedir.

Tablo 4.3: Model tahmin görüntüleri ve ground-truth.

Gauss FDA	GFOSMix	SFOSMix	Ground-Truth
			
			
			

Tablo 4.4: Model tahmin görüntüleri.

OM(UNet) FOSMix	FDA (B = 0.01)	FM+OM DCT FOSMix	FM+OM(UNet) DCT FOSMix
			
			
			

Tablo 3 ve 4’te geliştirilen yöntemlere ait model tahminlemelerine ve ground-truth değerleri verilmiştir. Bu görseller sayesinde model performansı otonom sürüş için önemli olan sınıflar göz ile anlaşılabilir. Örnek olarak araç için en önemli sınıflardan biri olan yolu en iyi bölütleyen model hangisidir gibi bir yaklaşım ile bu görseller incelenebilir.

Tablo 4.5: Geliştirilen yöntemlere ait metrik sonuçları.

Metrik	FDA	GFD A	Gauss FOSMi x	OM FOSMi x	SFOSMi x	FM+O M DCT FOSMi x	UNet FM+O M DCT FOSMi x
Piksel Doğruluğu	85.37 %	85.35 %	85.32%	80.43%	85.05%	74.87%	72.40%
Ortalama Sınıf Doğruluğu	56.82 %	51.39 %	53.40%	44.72%	54.89%	43.23%	41.78%
Frekans Ağırlıklı IoU	75.92 %	74.92 %	75.58%	69.49%	75.34%	62.10%	59.03%
Mean IoU	42.31 %	42.03 %	41.53%	34.10%	41.81%	31.10%	28.52%
Ortalama Dice	55.46 %	55.18 %	54.39%	46.09%	54.91%	42.73%	39.91%
Ortalama Keskinlik	60.78 %	67.80 %	62.79%	54.43%	64.18%	58.33%	58.68%
Ortalama Duyarlılık	56.82 %	51.39 %	53.40%	44.72%	54.89%	43.23%	41.78%
Ortalama F1 Skoru	55.46 %	55.18 %	54.39%	46.09%	54.91%	42.73%	39.91%

Tablo 5’te frekans tabanlı alan uyarlama yöntemlerine ait metrik sonuçları sunulmuştur. Bu bölümde, farklı alan uyarlama yöntemlerinin hangi açılardan farklılıkları değerlendirilmektedir.

Piksel doğruluğu metriği incelendiğinde, tüm alan uyarlama tekniklerinin genel olarak yüksek doğruluk oranları sergilediği görülmektedir. Bu durum, modelin piksel düzeyinde genel başarı sağladığını göstermektedir.

Sınıf bazında ortalama doğruluk açısından değerlendirildiğinde, özellikle düşük sıklıkta görülen sınıflarda sabit kare maskeye sahip FOSMix [5] yöntemi daha başarılı bir performans ortaya koymuştur.

Frekans ağırlıklı IoU metriği, görüntüde daha büyük alan kaplayan sınıflar dikkate alındığında, FDA [3] ve GFDA [4] yöntemlerinin bu tür sınıflarda daha iyi tahminler gerçekleştirdiğini göstermektedir.

Ortalama IoU (mIoU) metriği açısından en yüksek başarı Gauss FDA (GFDA) yöntemi ile elde edilmiştir. Bu bulgu, GFDA'nın sınıflar arası genel tahmin başarımında üstün olduğunu işaret etmektedir.

Ortalama Dice metriği ile mIoU arasında benzer bir eğilim izlenmiştir; bu da modellerin bölütleme kalitesi açısından benzer sıralamaları sunduğunu göstermektedir.

Ortalama keskinlik metriği, GFDA yönteminin pozitif tahminlerde diğer yöntemlere kıyasla daha doğru sınıflandırma yaptığını göstermektedir. Buna karşın, FDA ve FOSMix yaklaşımlarında sınıflar arasında aşırı tahmin eğilimi gözlemlenmiştir.

Gerçek pozitif sınıfları en iyi şekilde tahminleyen model ise sabit kare maskeli FOSMix yöntemi olmuştur.

Keskinlik ve duyarlılığın ortalamasını temsil eden ortalama F1 skoru açısından en başarılı model, kare maske kullanan alan uyarlama yöntemi olan FDA olarak belirlenmiştir.

FDA ile FOSMix yöntemlerinin bellek tüketimleri, yürütme süreleri gibi bu yöntemler kendi arasında karşılaştırılmıştır. Tablo 4.6 ve 4.7'de bu karşılaştırmalara yer verilmiştir.

Tablo 4.6: Yürütme süresi karşılaştırması.

Yöntem	Zaman (ms)	
	512x512 Görüntü	512x1024 Görüntü
FDA	1.15	1.41
GFDA	1.57	1.74
SFOSMix - FFT	1.24	1.44
GFOSMix – FFT	1.55	1.45
FOSMix Optimal Maske U-Net	14.42	27.50
FOSMix Optimal Maske Klasik	1.68	2.12

Tablo 4.6 incelendiğinde FDA Klasik yönteminin 1.15ms ile en hızlı yöntem olduğu gözükmemektedir. Burada elde edilen hız karşılaştırmaları model eğitimi esnasında avantajlı olmaktadır. Yüksek başarımlı gösteren modeller hızlıca eğitilip test alanında kullanıma açılabilir.

Tablo 4.7: Bellek tüketimi karşılaştırması.

Yöntem	Bellek Tüketimi (mB)	
	512x512 Görüntü	512x1024 Görüntü
FDA	54	88.3
GFDA	67	117.3
SFOSMix - FFT	54.5	90.3
GFOSMix – FFT	54.5	89.3
FOSMix Optimal Maske U-Net	1045	2065.3
FOSMix Optimal Maske Klasik	61.4	102.3

Şekil 4.6 incelendiğinde ise tez için kullanılan veri seti çözünürlüğü ile aynı olarak 512x1024 çözünürlüğe sahip bir görüntü için kare maske ile FDA 88.3MB boyutunda yer kaplaması ile RAM kullanımı açısından diğer yöntemlere göre düşük sistem gereksinimleri ile eğitim yapmayı kolaylaştırmaktadır.

5. SONUÇLAR VE YORUMLAR

Bu tez çalışmasında, frekans alanında farklı biçimlerde gerçekleştirilen alan uyarlama yöntemleri ele alınmıştır. Zorlu denetimsiz öğrenme senaryoları altında, frekans tabanlı alan uyarlama yaklaşımlarının anlamsal bölütleme üzerindeki performansları değerlendirilmiştir. GTA5 [1] video oyunu veri kümesi kaynak alan, Cityscapes [2] veri kümesi ise hedef alan olarak seçilmiş ve bu iki veri kümesi arasında dağılımsal farklılıkları azaltmak amacıyla çalışmalar yürütülmüştür. Böylece, sentetik ve gerçek veriler arasında bir bilgi köprüsü kurularak, alan kayması giderilmeye çalışılmıştır.

Literatürde önerilen Fourier Alan Uyarlaması (FDA) yönteminden esinlenilerek, Gauss FDA [4], Gauss FOSMix [5], Optimal Maske FOSMix FFT, Sabit Kare Maske FOSMix FFT, Optimal Maske FOSMix DCT ve Optimal UNet Maske FOSMix DCT olmak üzere 6 farklı frekans tabanlı alan uyarlama yöntemi test edilmiştir. Bu yöntemlere ait performans sonuçları 4. bölümde yer alan Tablo 4 ve Tablo 5'te detaylı olarak sunulmuştur.

Elde edilen sonuçlar incelendiğinde, Gauss FDA yönteminin mIoU, F1 skoru ve piksel doğruluğu gibi hem sınıf bazlı hem de ortalama bazlı metriklerde en yüksek başarıyı sağladığı görülmektedir. Özellikle Gauss penceresi kullanılarak frekans geçişinin yumuşatılması, stil transferi kapasitesini olumlu yönde etkilemiştir.

Sabit Kare Maske ile tasarlanan FOSMix yöntemi ise özellikle sınıf bazında tahmin dengesini ölçen duyarlılık ve ortalama sınıf doğruluğu metriklerinde başarılı sonuçlar vermiştir. Bu yöntem aynı zamanda mIoU açısından da rekabetçi bir performans ortaya koymuştur.

U-Net [30] tabanlı ve öğrenilebilir yapıda tasarlanan optimal maske ile eğitilen optimal karışım FOSMix yöntemi ise tüm metriklerde diğer yöntemlerin gerisinde kalmıştır. Eğitim stratejisinin sabit olması nedeniyle, modelin maske öğrenimi için yeterli adım sayısına ulaşamaması bu düşüşün başlıca nedenidir. Bu durum, öğrenilebilir maskelerin eğitim süresine duyarlı yapılar olduğunu ve daha gelişmiş optimizasyon stratejileri ile performanslarının artırılabilirliğini göstermektedir.

FOSMix aslında bir frekans karıştırma yöntemi olarak kullanıldığından dolayı düşük frekanslar ile ilgilenmeyi hedefleyen bir alan uyarlama yöntemi olduğu için optimal maske tabanlı FOSMix yaklaşımları bu tarz bir problemin çözümü için yeterince iyi performans gösteremedi.

Sonuç olarak, bu tez kapsamında üzerinde çalışılan frekans tabanlı alan uyarlama yöntemleri, denetimsiz öğrenme bağlamında gerçek dünya verilerindeki anlamsal bölütleme başarımını artırma potansiyeli taşımaktadır. Özellikle düşük etiketli ya da hiç etiket içermeyen hedef alanlar için, stil transferinde frekans bileşenlerinin etkili bir şekilde modellenmesi alan uyarlama süreçlerinde kritik rol oynamaktadır. Elde edilen bulgular, gelecekte farklı frekans pencereleri, öğrenilebilir maske yapıları ve dikkat mekanizmaları ile bu yöntemlerin daha da optimize edilebileceğini göstermektedir. Bu doğrultuda, frekans alanında geliştirilecek yeni yaklaşımların hem akademik çalışmalar hem de pratik uygulamalar için önemli katkılar sunması beklenmektedir.

6. GELECEK ÇALIŞMALAR

FOSMix [5] yöntemi için kullanılan optimal maskenin eğitiminin iyileştirilmesi ile stil aktarımını çok daha başarılı yapabilecek bir filtre tasarımı gerçekleştirilmek hedeflenmektedir. Maskeleme esnasında aynı zamanda anlamsal bölütleme için kullanılan veri kümesinde bulunan sınıflara duyarlı maskeler geliştirerek bu tarz çok sınıflı problemleri [8] çözümlemede arka planda kalan az sayıda olan sınıflar için faydalı olacaktır.

Tek bir alan (single-source) verisiyle sınırlı kalmak yerine, farklı alanlardan elde edilen çoklu veri kaynaklarının kullanılması planlanmaktadır. Bu yaklaşım, FOSMix algoritmasında stil aktarımının daha çeşitli ve zengin frekans bilgileri üzerinden yapılmasına olanak tanıyacak ve algoritmayı daha duyarlı hale getirecektir. Literatürde bu yaklaşım, çok kaynaklı alan uyarlama (multi-source domain adaptation) olarak adlandırılmaktadır.

Ayrıca, mevcut yöntemlerde yaygın olarak kullanılan düşük frekans aktarımına ek olarak, farklı frekans aralıklarını hedefleyen bant geçiren filtre tasarımlarının uygulanması amaçlanmaktadır. Böylelikle yalnızca genel stil bilgisi değil, farklı bantlarda taşınan yapısal ve anlamsal özelliklerin de uyarlanması mümkün olacaktır.

Son olarak, geliştirilecek bu frekans tabanlı alan uyarlama yöntemlerinin etkinliği, güncel ve güçlü yapay zeka modelleri üzerinde test edilecektir. Bu kapsamda özellikle son dönemde öne çıkan SegFormer [51] ve Swin Transformer [52] gibi transformer tabanlı anlamsal bölütleme modelleri tercih edilerek performans karşılaştırmaları yapılacaktır.

KAYNAKLAR

- [1] Richter S. R., Vineet V., Roth S., Koltun V., (2016), "Playing for Data: Ground Truth from Computer Games", European Conference on Computer Vision (ECCV), LNCS, vol. 9906, 102-118, Springer International Publishing.
- [2] Cordts M. vd., (2016), "The Cityscapes Dataset for Semantic Urban Scene Understanding", arXiv:1604.01685.
- [3] Yang Y., Soatto S., (2020), "FDA: Fourier Domain Adaptation for Semantic Segmentation", arXiv:2004.05498.
- [4] Liu X., Guo X., Liu Y., Yuan Y., (2021), "Consolidated domain adaptive detection and localization framework for cross-device colonoscopic images", Medical Image Analysis, 71, 102052.
- [5] Iizuka R., Xia J., Yokoya N., (2024), "Frequency-Based Optimal Style Mix for Domain Generalization in Semantic Segmentation of Remote Sensing Images", IEEE Transactions on Geoscience and Remote Sensing, 62, 1-14.
- [6] Krizhevsky A., Sutskever I., Hinton G. E., (2017), "ImageNet classification with deep convolutional neural networks", Communications of the ACM, 60(6), 84-90.
- [7] Long J., Shelhamer E., Darrell T., (2014), "Fully Convolutional Networks for Semantic Segmentation", arXiv:1411.4038.
- [8] Kumar V., Patil H., Lal R., Chakraborty A., (2023), "Improving domain adaptation through class aware frequency transformation", International Journal of Computer Vision, 131(11), 2888-2907.
- [9] Chen L. C., Papandreou G., Kokkinos I., Murphy K., Yuille A. L., (2016), "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs", arXiv:1606.00915.
- [10] Chen J., Wang C., Tong Y., (2019), "AtICNet: semantic segmentation with atrous spatial pyramid pooling in image cascade network", EURASIP Journal on Wireless Communications and Networking, 2019(1), 146.
- [11] LeCun Y., Bengio Y., Hinton G., (2015), "Deep learning", Nature, 521(7553), 436-444.
- [12] Zafar A. vd., (2022), "A Comparison of Pooling Methods for Convolutional Neural Networks", Applied Sciences, 12(17), 8643.
- [13] Srivastava N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdinov R., (2014), "Dropout: a simple way to prevent neural networks from overfitting", Journal of Machine Learning Research, 15(1), 1929-1958.

- [14]<https://ayyucekizrak.medium.com/derin-%C3%B6%C4%9Frenme-i%C3%A7in-aktivasyon-fonksiyonlar%C4%B1n%C4%B1n-kar%C5%9F%C4%B1la%C5%9Ft%C4%B1r%C4%B1lmas%C4%B1-cee17fd1d9cd>, (Erişim Tarihi: 21/06/2025).
- [15] Salimi M., Loni M., Afshar S., Cicchetti A., Sirjani M., (2024), "ConstScene: Dataset and Model for Advancing Robust Semantic Segmentation in Construction Environments", arXiv:2312.16516.
- [16] Zhou B., Zhao H., Puig X., Fidler S., Barriuso A., Torralba A., (2017), "Scene Parsing through ADE20K Dataset", 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 5122-5130, Honolulu, HI, USA, July.
- [17] Garcia-Garcia A., Orts-Escolano S., Oprea S., Villena-Martinez V., Garcia-Rodriguez J., (2017), "A Review on Deep Learning Techniques Applied to Semantic Segmentation", arXiv:1704.06857.
- [18] He K., Gkioxari G., Dollár P., Girshick R., (2017), "Mask R-CNN", arXiv:1703.06870.
- [19] Lowe D. G., (2004), "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision, 60(2), 91-110.
- [20] Dalal N., Triggs B., (2005), "Histograms of Oriented Gradients for Human Detection", 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 886-893, San Diego, CA, USA.
- [21] Girshick R., Donahue J., Darrell T., Malik J., (2016), "Region-Based Convolutional Networks for Accurate Object Detection and Segmentation", IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(1), 142-158.
- [22] Cortes C., Vapnik V., (1995), "Support-vector networks", Machine Learning, 20(3), 273-297.
- [23] Krizhevsky A., Sutskever I., Hinton G. E., (2012), "ImageNet Classification with Deep Convolutional Neural Networks", Advances in Neural Information Processing Systems, Curran Associates Inc.
- [24] Simonyan K., Zisserman A., (2014), "Very Deep Convolutional Networks for Large-Scale Image Recognition", arXiv:1409.1556.
- [25] Szegedy C. vd., (2014), "Going Deeper with Convolutions", arXiv:1409.4842.
- [26] He K., Zhang X., Ren S., Sun J., (2015), "Deep Residual Learning for Image Recognition", arXiv:1512.03385.
- [27] Badrinarayanan V., Kendall A., Cipolla R., (2015), "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation", arXiv:1511.00561.

- [28] Ronneberger O., Fischer P., Brox T., (2015), "U-Net: Convolutional Networks for Biomedical Image Segmentation", arXiv:1505.04597.
- [29] Lee J., Oh S. J., Yun S., Choe J., Kim E., Yoon S., (2022), "Weakly Supervised Semantic Segmentation using Out-of-Distribution Data", arXiv:2203.03860.
- [30] Yu Y. C., Lin H. T., (2023), "Semi-Supervised Domain Adaptation with Source Label Adaptation", arXiv:2302.02335.
- [31] Hoffman J. vd., (2017), "CyCADA: Cycle-Consistent Adversarial Domain Adaptation", arXiv:1711.03213.
- [32] Goodfellow I. J. vd., (2014), "Generative Adversarial Networks", arXiv:1406.2661.
- [33] Wang H., Shen T., Zhang W., Duan L., Mei T., (2020), "Classes Matter: A Fine-grained Adversarial Approach to Cross-domain Semantic Segmentation", arXiv:2007.09222.
- [34] Zhu J. Y., Park T., Isola P., Efros A. A., (2017), "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks", arXiv:1703.10593.
- [35] Nussbaumer H. J., (1981), "The Fast Fourier Transform", Fast Fourier Transform and Convolution Algorithms, Springer Series in Information Sciences, vol. 2, 80-111, Springer Berlin Heidelberg.
- [36] Basak H., Yin Z., (2023), "Semi-supervised Domain Adaptive Medical Image Segmentation through Consistency Regularized Disentangled Contrastive Learning", arXiv:2307.02798.
- [37] Zeng M., (2023), "GF2DA: Gaussian Filter Fourier Domain Adaptation", Proceedings of the 2023 International Conference on Frontiers of Artificial Intelligence and Machine Learning, 64-69, Beijing, China, April.
- [38] Ahmed N., Natarajan T., Rao K. R., (1974), "Discrete Cosine Transform", IEEE Transactions on Computers, C-23(1), 90-93.
- [39] Ros G., Sellart L., Materzynska J., Vazquez D., Lopez A. M., (2016), "The SYNTHIA Dataset: A Large Collection of Synthetic Images for Semantic Segmentation of Urban Scenes", 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 3234-3243, Las Vegas, NV, USA, June.
- [40] Everingham M., Van Gool L., Williams C. K. I., Winn J., Zisserman A., (2010), "The Pascal Visual Object Classes (VOC) Challenge", International Journal of Computer Vision, 88(2), 303-338.
- [41] Geiger A., Lenz P., Urtasun R., (2012), "Are we ready for autonomous driving? The KITTI vision benchmark suite", 2012 IEEE Conference on Computer Vision and Pattern Recognition, 3354-3361, Providence, RI, USA, June.

- [42] Zhou B. vd., (2016), "Semantic Understanding of Scenes through the ADE20K Dataset", arXiv:1608.05442.
- [43] Yu F. vd., (2020), "BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning", 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2633-2642, Seattle, WA, USA, June.
- [44] Brostow G. J., Fauqueur J., Cipolla R., (2009), "Semantic object classes in video: A high-definition ground truth database", Pattern Recognition Letters, 30(2), 88-97.
- [45]<https://numpy.org/doc/2.2/reference/generated/numpy.fft.rfft.html>, (Erişim Tarihi: 23/06/2025).
- [46] Ruder S., (2016), "An overview of gradient descent optimization algorithms", arXiv:1609.04747.
- [47] Kingma D. P., Ba J., (2014), "Adam: A Method for Stochastic Optimization", arXiv:1412.6980.
- [48]<https://www.sciencedirect.com/topics/computer-science/softmax-function>, (Erişim Tarihi: 25/06/2025).

ÖZGEÇMİŞ

Yakup Yıldız, 2017 yılında Özel İrfan Koleji Anadolu Lisesi'nden mezun olmuştur. 2022 yılında Lisans eğitimini Gebze Teknik Üniversitesi Elektronik Mühendisliği Bölümü'nde tamamlamıştır. 2023 yılından itibaren SmartUniversal firmasında Bilgisayarla Görü Mühendisi olarak kariyerine devam etmektedir.



TEZ METNİNDEN TÜRETİLEN YAYINLAR/SUNUMLAR

Yıldız, Y., Kayabol, K. (2025). Frekans Tabanlı Alan Uyarlama Yöntemlerinin Anlamsal Bölütleme Başarımlarının Karşılaştırılması. 9. GTÜ Lisansüstü Araştırmalar Sempozyumu, Gebze Teknik Üniversitesi, Gebze/Kocaeli, 26-27 Mayıs 2025, ss. 38-39

