

TEKNOFEST

HAVACILIK, UZAY VE TEKNOLOJİ FESTİVALİ

SAĞLIKTA YAPAY ZEKA YARIŞMASI

KRİTİK TASARIM VE ÖN BULGU RAPORU

ETUAI

#58322

TEKNOFEST
HAVACILIK, UZAY VE TEKNOLOJİ FESTİVALİ

İçindekiler

1. Kullanılması planlanan GÜNCEL donanımlar ve özellikleri (10 puan).....	3
1.1. Eğitim Donanımları (5 puan).....	3
1.2. Test Donanımları (5 puan).....	3
2. Çözüm için kullanılan yöntem ve mimari (30 puan)	4
2.1. ResNet tabanlı sınıflandırma modelleri	4
2.2. Segmentasyon modelleri.....	4
2.3. Kullanılan Hata Fonksiyonları.....	6
2.3.1. Focal Loss	6
2.3.2. IoU Loss	6
2.3.3. Dice Loss.....	7
3. Deney ve eğitim aşamalarında kullanılan veri setleri (30 puan).....	7
3.1. Ön-işlem	8
3.1.1. Hounsfield Birim Dönüşümü	8
3.1.2. Normalizasyon	8
3.1.3. Sıfır-ortalama	9
3.1.4. Yeniden boyutlandırma	9
3.2. Veri Arttırımı ve Dönüşümü.....	9
4. Elde edilen deneysel sonuçların sunumu (30 puan).....	10
4.1. Sınıflandırma sonuçları.....	10
4.2. Segmentasyon Sonuçları.....	11
5. Referanslar	13

1. Kullanılması planlanan GÜNCEL donanımlar ve özellikleri (10 puan)

1.1. Eğitim Donanımları (5 puan)

Eğitimler için 64 bit işletim sisteminde 3.30 GHz Intel(R) Xeon(R) W-2155 CPU işlemcili bir sistem kullanılmıştır. Sistemin yüklü belleği (RAM) 128 GB'tır. RAM değerinin yüksek olması veri setinin yazılım platformuna yüklenebilmesi ve işlenebilmesi için oldukça önem taşımaktadır. Özellikle açık kaynaklı veri setleri de kullanılabileceğinden ve halihazırda bulunan CT görüntüleri yüksek çözünürlükte olduğundan yüksek bellek oldukça kritiktir. Derin öğrenme modelleri eğitilirken Grafik İşlemci Ünitesi (GPU) kullanılması neredeyse zorunluluktur. Özellikle hız açısından GPU kullanımı ciddi avantaj sağlamaktadır. Bu sebeple model eğitiminde yüksek hızda eğitim sağlayabilen CUDA destekli iki adet GDDR6 12 GB NVIDIA GeForce RTX 3060 ekran kartı kullanılmıştır. Derin öğrenme modelleri eğitilirken parti boyutu (Batch size) eğitimin hızlı tamamlanması için oldukça önemli bir parametre olmakla birlikte küçüldükçe batch-norm gibi katmanlardan dolayı eğitim performansında sorunlar ortaya çıkabilmektedir. CT görüntülerinin çözünürlüklerinin yüksek olduğu düşünüldüğünde, 12 GB hafızalı iki adet ekran kartı kullanılması, eğitimlerde parti boyutunun eğitim performansını düşürmeden seçilebilmesini sağlamaktadır. Örneğin görüntü çözünürlüğünün 512×512 olduğu durumda ve ImageNet ile eğitilmiş bir modelden transfer learning yapıldığı düşünüldüğünde Pytorch için girdi boyutu (parti boyutu $\times 3 \times 512 \times 512$) olacaktır. Yaptığımız hesaplara göre bu girdi boyutlarında 2 adet 12 GB belirtilen ekran kartının kullanılması, 10 değerindeki parti boyutunun kullanılmasına izin verecektir. Bu parti boyutu hiper-parametreler doğru seçildiği takdirde (örn: öğrenme oranı, normalizasyon yöntemi vs) yüksek performansta eğitim sağlayabilecektir.

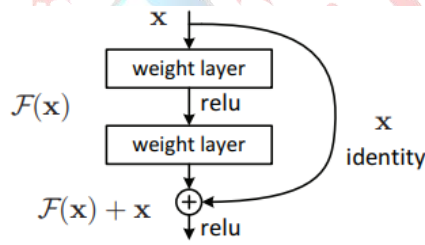
1.2. Test Donanımları (5 puan)

Test kısımları için de aynı donanımlar kullanılacaktır. Fakat test yaparken parti boyutu normalin 3 katına çıkarılabilmektedir. Bunun sebebi sadece model ağırlıklarının tutulup, gradyan ve momentum bilgilerinin tutulmasına gerek olmamasındandır. Aynı şekilde yalnızca ileri besleme yapılacağından, eğitim performansından daha hızlı bir sonuç alınabilmektedir.

2. Çözüm için kullanılan yöntem ve mimari (30 puan)

2.1. ResNet tabanlı sınıflandırma modelleri

Yarışmanın ilk aşaması olan sınıflandırma problemi için deneysel olarak incelenen ve önerilen model mimarisi ResNet [1] olarak belirlenmiştir. Her ne kadar sınıflandırma problemlerinde VGG, DenseNet gibi mimariler de yüksek başarı sergilese de ResNet mimarileri özellikle çok derin modellerde ortaya çıkabilen kaybolan gradyanlar problemine residual bağlantıları kullanarak çözüm sağlayabilmektedir. ImageNet [2] sınıflandırma probleminde gösterdiği yüksek başarı sayesinde günümüzde birçok modern derin öğrenme problemlerinde sıkça kullanılmaktadır. ResNet kendi içerisinde de bulundurduğu toplam katman sayısına göre aşağıdaki gibi farklı modellere ayrılabilir: ResNet-18, ResNet-34, ResNet-50, ResNet-101. Şekil 1’de ResNet mimarilerinde kullanılan residual blok gösterilmiştir.



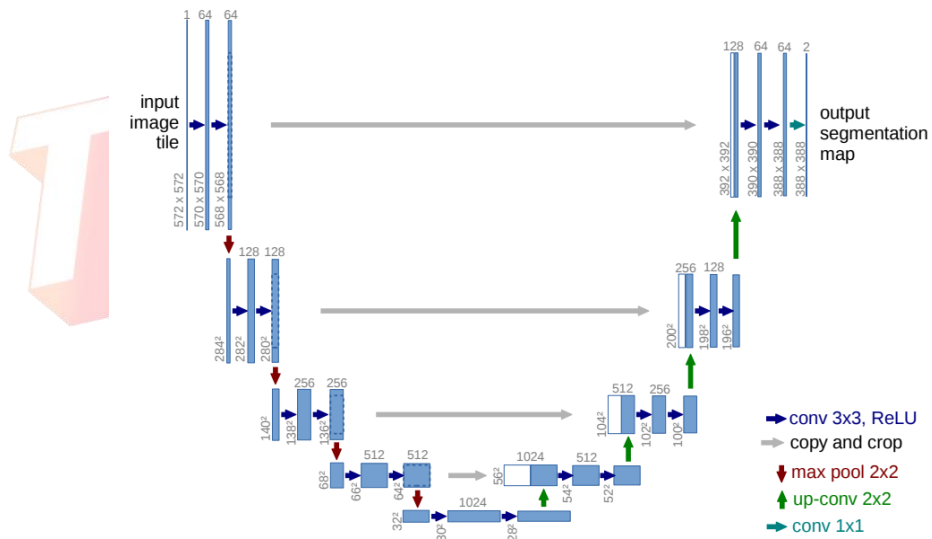
Şekil 1. Residual blok [1]

Yarışmanın ilk aşaması kapsamında da yukarıdaki belirtilen katman sayılarına göre model mimarileri eğitilmiş, incelenmiş ve deneysel sonuçlar elde edilmiştir.

2.2. Segmentasyon modelleri

Biyomedikal görüntüler üzerinden derin öğrenme modelleri kullanarak yüksek başarılı segmentasyon performansı elde etmek oldukça zor bir süreçtir. Bunun temel sebebi, derin öğrenme modellerinin yüksek performansta eğitim için yüz binlere hatta milyonlarca veriye ihtiyaç duymasındandır. Biyomedikal görüntüleme gibi problemlerde bu kadar veri toplayabilmek pratikte mümkün olmadığından, araştırmacılar az veri kullanılarak yüksek performanslı eğitim sonucu elde edebilmek adına, eğitim performansını arttıracak çeşitli metotlar önermişlerdir. Örneğin UNet [3] mimarisi, Konvolüsyonel Enkoder Dekoder (KED) tabanlı bir mimari olup, enkoder bloğundaki düşük seviyeli özellik matrisleri ile dekoder kısmında bulunan yüksek dereceli özellik matrisleri arasında birleştirme işlemi kullanarak standard KED mimarilerine göre eğitim performansında ciddi bir artış gözlemlemişlerdir. U-Net mimarisi uzun bir süre SotA (state-of-the-art) modellerden birisi olarak kullanılsa da günümüz GPU’larının ve donanımsal elemanların gelişmesi ile yerini daha güçlü, derin mimarilere bırakmıştır. Örneğin geçtiğimiz yıllarda önerilen ResUNet [4] mimarisi ResNet

modellerinin yapı taşı olan residual bağlantıları UNet ile birleştirerek ve mimari derinleştirilerek segmentasyon problemlerinde yüksek performans elde etmeyi başarmıştır. ResUNet mimarisini takiben önerilen ve ResUNet temel alınarak içerisinde çeşitli geliştirmeler bulunduran ResUNet++ [5] mimarisi, bilgisayar destekli polyp segmentasyonu probleminde yüksek başarı sergilemeyi başarmıştır. ResUNet++ mimarisi ResUNet'e ek olarak içerisinde squeeze-excite (SE) [6], Atrous Spatial Pyramid Pooling (ASPP) [7] ve Attention [8, 9] olmak üzere üç temel blok bulundurmaktadır. SE blokları KED mimarisinin enkoder kısmında kullanılmaktadır. SE bloklarındaki temel amaç konvolüsyon operasyonu ile elde edilen özellik cevaplarını yeniden kalibre ederek birbiri ile ilişkili özellik matrislerini çıkarabilmektir. SE blokları bu şekilde modelin önemli özellik matrislerine olan hassaslığını artırır ve önemli olmayan özellik matrislerinin katkısını azaltır. ASPP blokları ise enkoder ve dekoder blokları arası geçişte özellik matrislerinin farklı bölgelerindeki (multi-scale) bilgilerin beraber işlenmesi amacıyla kullanılmaktadır. 4 farklı konvolüsyonel operasyon paralel bir şekilde ve farklı geciktirme (dilation) oranları ile uygulanır ve en son çıkışta birleştirilir. Attention blokları ise dekoder kısmında kullanılmaktadır. Enkoder kısmından gelen düşük seviye ve dekoderden gelen yüksek seviye özellik matrislerini beraber işleyerek özellik matrisleri içerisinde bulunan ve segmentasyon haritasının oluşmasını sağlayan önemli bölgelere ilgiyi artırarak (attention) segmentasyon performansının artırılmasını sağlamaktadır.



Şekil 2. UNet mimarisi [3]

Segmentasyon kısmında değerlendirilen son model tipi ise R2-UNet [10] mimarisine Attention blok eklenmesi ile oluşturulan R2-Att-UNet mimarisidir. R2-Att-UNet, ResUNet mimarisindeki gibi residual bağlantıları kullanmakla birlikte her bir enkoder bloğu sonrası öz yinelemeli bağlantılar kullanılarak (kapalı döngü) özellik matrislerinin işleme performansı

arttırılmaktadır. ResUNet++ mimarisine benzer olarak R2-Att-UNet mimarisinde de Attention blokları kullanılarak dekoder bloğunun segmentasyon karakteristiğini belirleme performansı arttırılmaktadır.

Yukarıda bahsedilen segmentasyon modellerinin tamamında, temelde konvolüsyonel operasyon ve lineer olmayan aktivasyon fonksiyonları arasında Batch Normalization (BN) [11] kullanılmaktadır. Her ne kadar BN katmanları aşırı öğrenmeyi engellemek, eğitimi hızlandırmak gibi özelliklere sahip olsa da, eğitim sırasında bir partide kullanılan resimler boyunca istatistiksel özellik saptama temelli normalizasyon işlemi yaptığından ve özellikle küçük parti boyutlarında parti resimleri istatistiksel özelliklerinin yeterince genel saptanamaması sonucunda, eğitim performansında stabil olmayan durumlar gerçekleşebilmektedir. Küçük parti boyutları için (örneğin parti boyutu < 32) BN'e alternatif olarak önerilen ve basit fakat etkili bir normalizasyon metodu olan Group Normalization (GN) [12] küçük parti boyutları için BN'e göre daha iyi performans göstermesi sebebiyle bütün modellerde BN katmanları yerine kullanılmıştır.

2.3. Kullanılan Hata Fonksiyonları

Çalışmadan kullanılan hata fonksiyonları aşağıda verilmiştir.

2.3.1. Focal Loss

Segmentasyon problemlerinde performansı oldukça arttıran focal hata fonksiyonu (Denklem 1), kullandığı γ parametresi ile CE hatası ile öğrenilmesi zor örnekler ağırlık verebilmektedir. γ parametresi arttırıldığında eğitilen model yanlış bildiği örneklerde daha fazla cezalandırılmaktadır.

$$FL(p_t) = -(1 - p_t)^\gamma \log(p_t) \quad 1$$

2.3.2. IoU Loss

Kesişim-bölü-birleşim, intersection-over-union, (IoU) hata fonksiyonu tam olarak adından anlaşıldığı ifadeyi gerçekler. Tahmin edilen segmentasyon haritası ile gerçek değer ne kadar üst üste binerse IoU ifadesi o kadar büyür, haliyle 1 eksili hali de o kadar küçülür. Denklem 2'de bu fonksiyonun aşamaları görülmektedir.

$$IoU = \frac{I(X)}{U(X)}$$

$$I(X) = \sum_{v \in V} X_v * Y_v$$

$$U(X) = \sum_{v \in V} (X_v + Y_v - X_v * Y_v)$$

$$L_{IoU} = 1 - IoU \quad 2$$

2.3.3. Dice Loss

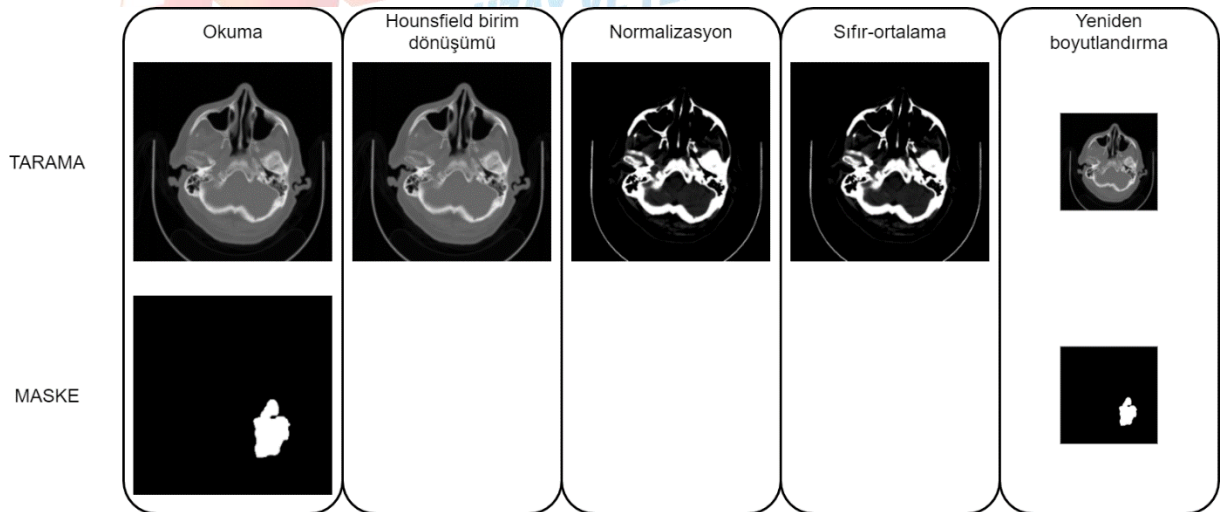
Dice hata fonksiyonu yapısı itibari ile IoU hata fonksiyonuna oldukça benzemektedir. Tam olarak, $\frac{2 \cdot IoU}{IoU + 1}$ olarak ifade edilebilir. Denklem 3'te iki sınıflı bir problem için olan matematiksel ifadesine yer verilmiştir.

$$DL_2 = 1 - \frac{\sum_{n=1}^N p_n r_n + \epsilon}{\sum_{n=1}^N p_n + r_n + \epsilon} - \frac{\sum_{n=1}^N (1 - p_n)(1 - r_n) + \epsilon}{\sum_{n=1}^N 2 - p_n - r_n + \epsilon} \quad 3$$

3. Deney ve eğitim aşamalarında kullanılan veri setleri (30 puan)

Sağlık Bakanlığı verilerine bakıldığında, 4427 adet inmeyok, 1130 adet iskemi, 1093 adet kanama verisi görülmektedir. Sınıflandırma probleminde kullanılacak hata fonksiyonuna göre bu veri dağılımları dikkate alınmalıdır.

Veriler PNG ve DICOM formatındadır. PNG formatı 0-255 arasında uint8 formatında verileri barındırırken, DICOM formatı float verileri de tutabilecek, medikal alanda artık standart haline gelmiş bir formattır. Her ne kadar görselleştirirken PNG formatı ile çalışmak kolaylık sağlasa da, ön işleme işlemleri ve model eğitimi esnasında DICOM verileri daha büyük öneme ve esnekliğe sahiptir. Ayrıca DICOM verileri, taramayı yapan makinenin özelliklerini, piksel arası aralığın gerçek hayatta nasıl bir fiziksel büyüklüğe denk geldiği gibi bilgileri de barındırır. Bu nedenle DICOM verileri, belirli ön işleme adımlarından geçirilerek, modelin eğitim süreci kolaylaştırılabilir ve hızlandırılabilir.



Şekil 3. Ön-işlem aşamaları

3.1. Ön-işlem

Makine öğrenmesi modelleri kullanılırken çok daha önemli olan ön-işlem aşaması, derin öğrenme modelleri kullanıldığında da mümkün olduğunca kullanılmalıdır ki model daha hızlı ve gürültüden uzak bir eğitim süreci geçirsin. Bu çalışmada ön-işlem aşamaları, Şekil 3’de görüldüğü, hounsfield birim dönüşümü, normalizasyon, sıfır-ortalama ve yeniden boyutlandırma aşamalarından oluşmaktadır.

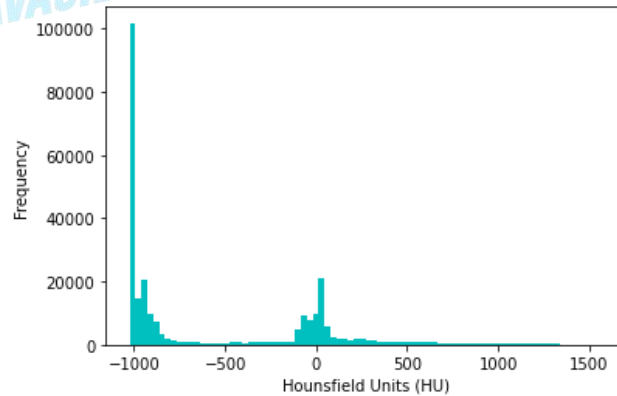
3.1.1. Hounsfield Birim Dönüşümü

DICOM verileri okunduğunda her bir pikseldeki atenüasyon katsayısına ulaşılır. Bu değer Denklem 4’de verilen dönüşüm ile Hounsfield birimine dönüştürülür. CT tarayıcıları bu birimi doğru ölçecek şekilde kalibre edildikleri için böyle bir dönüşümün yapılması veriseti içerisindeki istatistiksel uyumu arttıracaktır. Bu işlemi yapmadan önce silindir CT tarayıcıların varsayılan olarak atadığı -2000 değeri de hava olarak ayarlanmıştır.

$$HU = 1000 \times \frac{\mu - \mu_{su}}{\mu_{su} - \mu_{hava}} \quad 4$$

3.1.2. Normalizasyon

HU değerleri yaklaşık olarak -1000 ile +2000 arasında değişmektedir. Şekil 4’de örnek bir veri için HU değerlerinin dağılım eğrisi görülmektedir. -1000 civarlarında büyük bir yığılım ve 0 çevresinde bir yığılım gözlemlenmektedir. HU değerlerinin bulunduğu aralıklar çeşitli doku gruplarını ifade etmektedir. Tablo 1’e bakılacak olunursa, yumuşak doku, gri ve beyaz madde aralıklarının 0 ile 300 arasında olduğu görülmektedir. 0 HU değerinden düşük değerler ve 300 HU değerinden büyük değerler, problemin çözümü için gereksiz dokuları barındırmaktadır. Bu nedenle yalnızca belirtilen aralıktaki veriler korunup, hemen sonrasında da 0 ile 1 arasına normalize edilmişlerdir.



Şekil 4. Örnek bir veri için HU değerlerinin dağılım eğrisi

Tablo 1. Maddeye karşılık gelen Hounsfield değerleri

Madde	HU
Hava	-1000
Akciğer	-500
Yağ	-100 -- -50
Su	0
CSF	+15
Böbrek	+30
Kan	+30 -- +45
Kas	+10 -- +40
Gri madde	+37 -- +45
Beyaz madde	+20 -- +30
Karaciğer	+40 -- +60
Yumuşak doku	+100 -- +300
Kemik	+700 -- +3000

3.1.3. Sıfır-ortalama

Normalizasyon işleminden sonra eğitim verisinin ortalama değerine bakıldığında 0.102 olduğu görülmüştür. Bu nedenle tüm veriden 0.102 global ortalama değeri çıkarılmış ve veri dağılımı 0 ortalama çevresine alınmıştır.

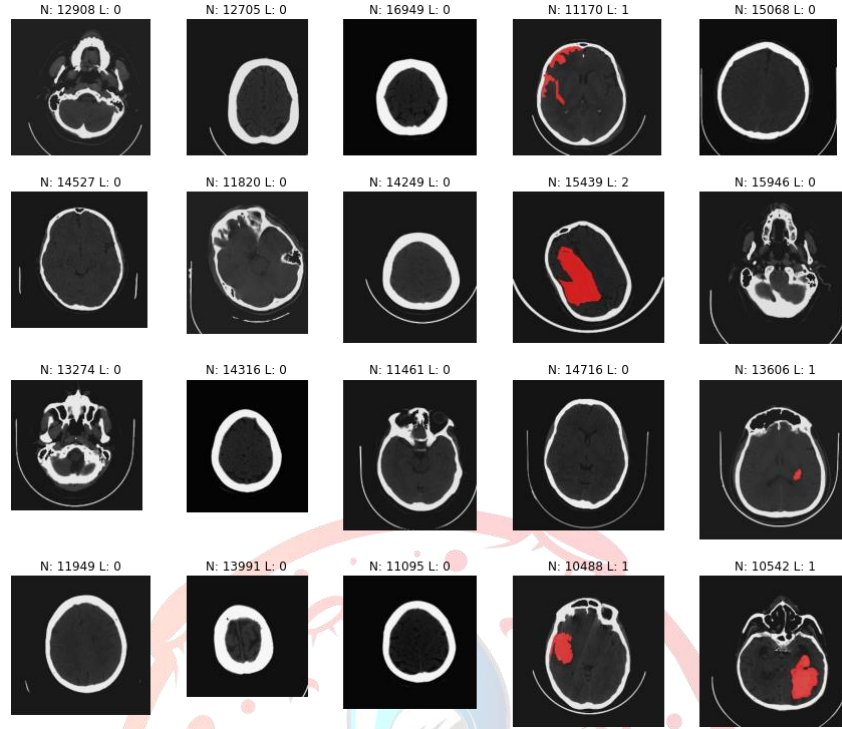
3.1.4. Yeniden boyutlandırma

DICOM verileri CT tarayacının özelliklerini de tutmaktadır. Bu özellikler arasında iki piksel arası uzaklığın fiziksel hayatta kaç mm'ye denk geldiği bilgisi de bulunmaktadır. Veriler incelendiğinde tüm resimler yaklaşık 512x512 olmasına rağmen, iki piksel aralığı oldukça farklılık göstermektedir. Model eğitimi esnasında fiziksel yapının daha iyi anlaşılması için bu değerlerin aynı dilden konuşması yani yeniden boyutlandırma yapılması gerekmektedir. Bundan dolayı fiziksel dünyada 1 mm olacak şekilde resimle interpolasyon uygulanarak yeniden boyutlandırılmışlardır.

Ön-işlem aşamalarından sonra 4'e 5lik örnek ızgara görüntüsü Şekil 5'te verilmiştir.

3.2. Veri Arttırımı ve Dönüşümü

Verileri çeşitlendirmek ve modelin ezberlemek yerine genel yapıyı anlamasını sağlamak için çeşitli arttırım (augmentation) ve dönüşüm (transformation) yöntemleri kullanılmıştır. Bunlar, ShiftScaleRotate (resimde sağa sola kaydırma), HorizontalFlip (yatay çevrim), VerticalFlip (dikey çevrim), ve .çeşitli gürültü ekleyen ya da görüntüyü bir miktar bozan dönüşümler (GridDistortion, ElasticTransform, OpticalDistortion, GridDropout, GaussianBlur, GlassBlur, ColorJitter, GaussNoise) uygulanıp, resimler aynı boyuta gelebilsin diye de CenterCrop (ortadan kesim) uygulanmıştır.



Şekil 5. Ön-işlem aşaması sonrası örnek görüntüler. Başlıklardaki N örneğin IDsini, L ise hangi sınıfa ait olduğunu göstermektedir. Kırmızı olarak işaretlenen yerler ise iskemi ya da kanama bölgelerini göstermektedir.

4. Elde edilen deneysel sonuçların sunumu (30 puan)

Bütün deneylerde, veriseti önce karıştırılmış, daha sonra %75 eğitim, %25 test olacak şekilde ikiye bölünmüştür. Ön-işlem aşamalarında gereken kısımlarda yalnızca eğitim verisi kullanılmıştır. Sonuç tablolarında ve grafiklerinde gösterilen değerler, test verisi üzerinden hesaplanmıştır.

4.1. Sınıflandırma sonuçları

Sınıflandırma model eğitimi Tablo 2. verilen hiper-parametreler ile yapılmış olup, 4 farklı ResNet mimarisine sahip model eğitilmiştir.

Tablo 3'ten de görüldüğü üzere model sonuçları ResNet34 mimarisi dışında oldukça yakın çıkmıştır. Uygun hiper-parametreler ayarlandığında ResNet34 ile de aynı başarımda sonuçlar alınacağı düşünülmektedir. Şuanki sonuçlara bakıldığında ResNet50 mimarisi en iyi sonuca sahiptir. Model karmaşıklığı arttığında sonucun iyileşmesi gerekir gibi düşünülebilir fakat yeterinde veri ile model beslenemezse, parametre sayısının artması, hatanın minimize olacağı uygun yeri bulmayı zorlaştırır. Bu nedenle model karmaşıklığı artmış olmasına rağmen performansta bir iyileşme görüşmemiştir.

Tablo 3 100 epoch sonundaki sonuçları gösterirken, doğruluk ve μ_{metric} grafikleri için ve Şekil 6 incelenebilir. Bu grafikteki μ_{metric} , duyarlılık ve özgüllüğün ortalaması olarak ifade edilmiştir.

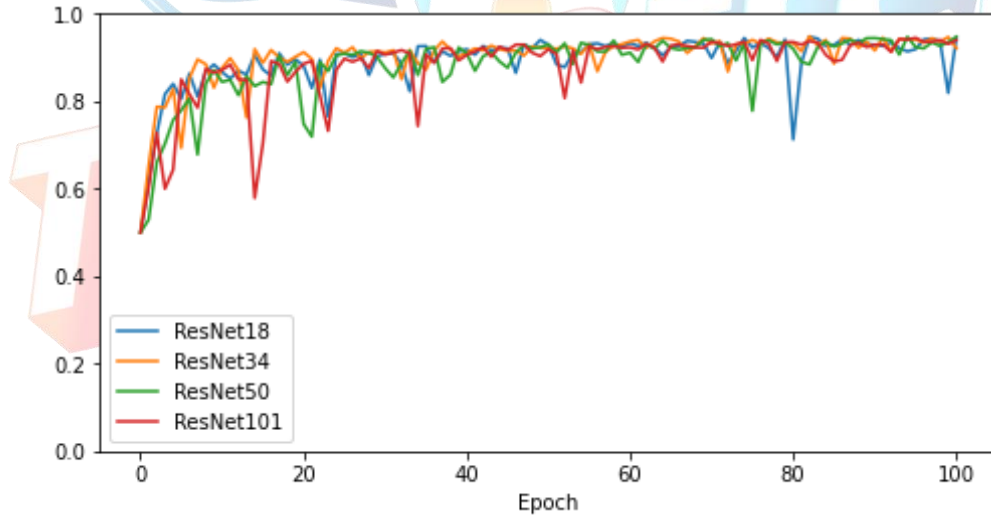
Bu grafiklere bakıldığında eğitim sırasında sıçramalar olduğu görülmektedir. Yığın sayısı artırılabilirse ve data augmentation daha uygun kullanılırsa bunun da çözüleceğini düşünmekteyiz.

Tablo 2. Sınıflandırma hiper-parametreleri

LR	1e-4
Epoch	100
Batch size	16
Optimizer	ADAM
Hata Fonk.	Cross-Entropy
Hata Fonk. Sınıf Ağırlıkları	[0.33, 0.67]

Tablo 3. Sınıflandırma sonuçları

	CE	Doğruluk (Accuracy)	μ_{metric}	Duyarlılık (Recall)	Özgüllük (Specificity)
ResNet18	0.1894	0.9464	0.9449	0.9407	0.9492
ResNet34	0.2101	0.9085	0.9212	0.9574	0.8851
ResNet50	0.1743	0.9549	0.9473	0.9259	0.9688
ResNet101	0.1747	0.9434	0.9398	0.9296	0.9501



Şekil 6. Sınıflandırma modelleri μ_{metric} sonuçları

4.2. Segmentasyon Sonuçları

Segmentasyon deneyleri Tablo 4'te verilen hiper-parametreler ile yapılmıştır. Bu tabloda tüm modeller için ortak olan parametreler verilmiş olup, deneyden deneye değişen parametreler sonuç tablosunda ifade edilmiştir. Deneyler

Tablo 5'te görüldüğü şekilde 5 farklı model ile yapılmıştır. İncelenen modeller arasında en iyi sonucu ResUNet++ isimli, Unet benzeri mimariye sahip olan fakat enkoder kısmında ResNet

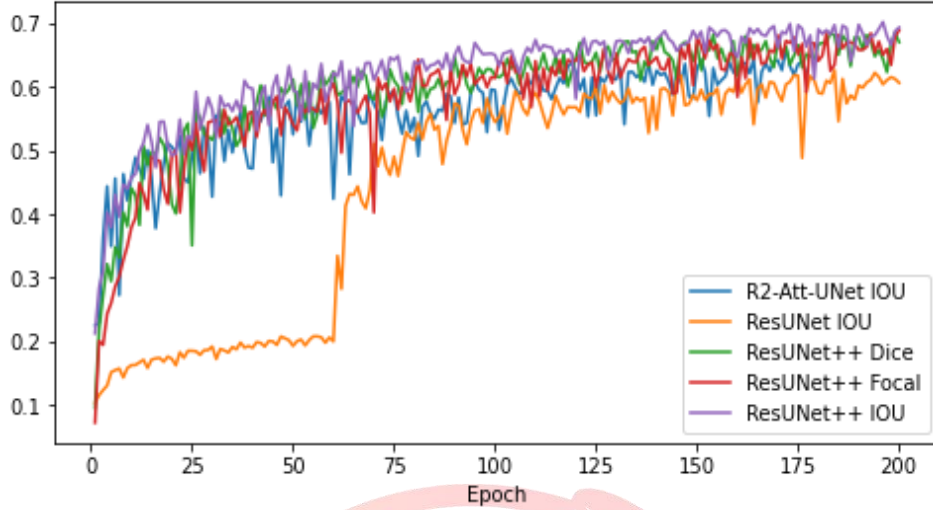
mimarisinden faydalanan, ayrıca içerisinde daha önceden bahsedilmiş olan SotA yapılan barındıran mimari vermiştir. Dice ve IoU hata fonksiyonları arasında pek bir farkın olmadığı fakat Focal hata fonksiyonunun biraz geride kaldığı görülmüştür. Dice ve IoU arasındaki fark daha önceden söylendiği gibi minimal olduğu için böyle bir sonuç çıkması gayet uygundur. Tabloda verilmiş olan DEIoU metrik değerleri ise yarışmada hesaplanacak olan metriktir. Burada uygulanan dilation ve erosion işlemleri, bu metriği orijinal IoU'dan biraz farklılaştırmış ve IoU hata fonksiyonunu optimize etmeye çalışan ResUNet++'ı öne çıkarmıştır. Şekil 7'de 200 epoch boyunca hataların nasıl hareket ettiği görülmektedir. Bu şekilde dikkatlice bakıldığında ResUNet ile ResUNet++ arasında nasıl bir başarımlık farkı olduğu görülmektedir. 60. epoch'a kadar ResUNet öğrenme gücü çökerken diğer modeller stabil bir şekilde öğrenmeyi sürdürmüşlerdir.

Tablo 4. Segmentasyon hiper-parametreleri

LR	1e-4
Epoch	200
Batch size	5
Optimizer	ADAM

Tablo 5. Segmentasyon sonuçları. IoU, kesişim bölü birleşim; DEIoU, Dilation ve Erosion işlemi yapılan IoU; μ_{metric} , yarışmada sınıflandırma kısmında verilen metriği ifade etmektedir. k katsayısı ise mimarileri katman boyutlarını belirleyen, büyüdükçe modeli genişleten bir değerdir.

	Hata Fonk.	k	Hata	IoU	DEIoU	μ_{metric}
R2AttUnet	IOU	0.25	0.4611	0.4611	0.6232	0.9167
ResUnet	IOU	1.00	0.4705	0.5295	0.6064	0.8984
ResUNet++	Dice	0.5	0.4173	0.5479	0.6704	0.9165
ResUNet++	Focal	0.5	0.0046	0.5371	0.6891	0.8836
ResUNet++	IOU	0.5	0.4541	0.5456	0.6932	0.9055



Şekil 7. Dilation ve Erosion işlemi yapılmış IoU (DEIoU) sonuçları

5. Referanslar

- [1] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016.
- [2] O. Russakovsky *et al.*, “ImageNet Large Scale Visual Recognition Challenge,” *Int. J. Comput. Vis.*, 2015.
- [3] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical image computing and computer-assisted intervention*, 2015, pp. 234–241.
- [4] Z. Zhang, Q. Liu, and Y. Wang, “Road Extraction by Deep Residual U-Net,” *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 749–753, May 2018.
- [5] D. Jha *et al.*, “ResUNet++: An Advanced Architecture for Medical Image Segmentation,” *Proc. - 2019 IEEE Int. Symp. Multimedia, ISM 2019*, pp. 225–230, Nov. 2019.
- [6] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, “Squeeze-and-Excitation Networks,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2011–2023, Sep. 2017.
- [7] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.
- [8] A. Vaswani *et al.*, “Attention is all you need,” in *Advances in Neural Information Processing Systems*, 2017, vol. 2017-December, pp. 5999–6009.
- [9] H. Li, P. Xiong, J. An, and L. Wang, “Pyramid Attention Network for Semantic Segmentation,” *Br. Mach. Vis. Conf. 2018, BMVC 2018*, May 2018.
- [10] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, “Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation,” Feb. 2018.
- [11] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” in *32nd International Conference on Machine Learning, ICML 2015*, 2015, vol. 1, pp. 448–456.
- [12] Y. Wu and K. He, “Group Normalization,” *Int. J. Comput. Vis.*, 2020.