

Tooth Segmentation using Transformers

Allahverdi Hajiye

Electronic and Computer Engineering

Özet

Bu rapor, panoramik ağız röntgeni görüntülerinden diş segmentasyonu yapmak için bir Vision Transformer (ViT) modelinin geliştirilmesi ve değerlendirilmesine yönelik kapsamlı bir çalışmayı sunmaktadır. Modelin performansını artırmak için tüm veri artırma teknikleri uygulanmış, özelleştirilmiş bir ViT modeli ve hibrit bir kayıp fonksiyonu kullanılmıştır. Sonuçlar, Intersection over Union (IoU), Dice katsayısı ve doğruluk gibi umut verici metriklerle modelin doğru segmentasyon yapma yeteneğini göstermektedir. Bu çalışma, kullanılan metodolojilere, karşılaşılan zorluklara ve yapılan iyileştirmelere vurgu yaparak diş hekimliği teşhislerinde sağlam bir model elde edilmesini sağlamıştır. Raporda ayrıca modelin mimarisinin ayrıntıları, kullanılan spesifik veri artırma teknikleri ve modelin eğitimi ve değerlendirilmesi süreci detaylı bir şekilde incelenmektedir.

Literatür İncelemesi

Diş segmentasyonu, diş hekimliği teşhisinde ve tedavi planlamasında kritik bir rol oynar. Bu alandaki doğru ve etkili segmentasyon yöntemleri, diş hekimlerine hastaların ağız sağlıklarını değerlendirme ve tedavi süreçlerini planlama konusunda önemli bir destek sağlar. Geleneksel görüntü işleme teknikleri, diş segmentasyonu için genellikle yetersiz kalmakta ve bu durum, daha gelişmiş algoritmaların ve modellerin geliştirilmesini gerektirmektedir.

Son yıllarda, derin öğrenme alanındaki ilerlemeler, tıbbi görüntüleme ve segmentasyon alanında çığır açıcı yenilikler sunmuştur. Bu bağlamda, Konvolüsyonel Sinir Ağları (CNN) ve U-Net gibi derin öğrenme mimarileri, tıbbi görüntü segmentasyonunda yaygın olarak kullanılmıştır. Ronneberger ve ark. (2015) tarafından tanıtılan U-Net modeli, özellikle biyomedikal görüntü

segmentasyonunda yüksek başarı sağlamış ve birçok çalışmada standart yöntem olarak benimsenmiştir.

Ancak, bu geleneksel derin öğrenme yöntemleri, özellikle uzun menzilli bağımlılıkları ve küresel bağlamı yakalamada sınırlamalara sahiptir. Transformer mimarileri, ilk olarak doğal dil işleme (NLP) alanında kullanılmak üzere geliştirilmiş olsa da, görüntü işleme görevlerinde de büyük potansiyel göstermektedir. Vaswani ve ark. (2017) tarafından tanıtılan "Attention is All You Need" çalışması, transformer modellerinin dikkat mekanizmaları sayesinde uzun menzilli bağımlılıkları yakalama yeteneğini vurgulamıştır.

Dosovitskiy ve ark. (2020) tarafından tanıtılan Vision Transformer (ViT) modeli, transformer mimarisinin görüntü tanıma görevlerinde de etkili olduğunu göstermiştir. ViT, görüntüleri sabit boyutlu yama (patch) dizilerine dönüştürerek bu yamalar üzerinde dikkat mekanizmaları uygulamaktadır. Bu yöntem, geleneksel CNN'lere kıyasla daha iyi performans göstermekte ve uzun menzilli bağımlılıkları daha etkili bir şekilde yakalayabilmektedir.

ViT modelinin diş segmentasyonu gibi tıbbi görüntü segmentasyonu görevlerine uygulanması, bu modelin uzun menzilli bağımlılıkları yakalama yeteneğinden faydalanarak daha doğru ve güvenilir segmentasyon sonuçları elde etmeyi amaçlamaktadır. Diğer yandan, ViT modelinin tek başına kullanımı, özellikle ince detayların segmentasyonunda yeterli olmayabilir. Bu nedenle, ViT modelinin konvolüsyonel katmanlarla birleştirilmesi, hem küresel bağlamın hem de yerel detayların daha iyi yakalanmasını sağlar.

Çalışmamızda, ViT modeli, diş segmentasyonu görevini gerçekleştirmek üzere konvolüsyonel katmanlarla birleştirilerek kullanılmıştır. Bu yaklaşım, hem transformer modellerinin uzun menzilli bağımlılıkları yakalama yeteneğini hem de konvolüsyonel katmanların yerel detayları yakalama yeteneğini bir araya getirerek daha etkili bir segmentasyon modeli oluşturmayı hedeflemektedir. Ayrıca, veri artırma teknikleri ve hibrit kayıp fonksiyonları gibi yöntemler de modelin performansını artırmak için kullanılmıştır.

Sonuç olarak, ViT modeline dayalı diş segmentasyonu çalışması, tıbbi görüntü segmentasyonu alanında önemli bir yenilik sunmakta ve gelecekte bu alanda yapılacak çalışmalara yol gösterici olmaktadır.

Giriş

Panoramik ağız röntgeni görüntülerinden diş segmentasyonu, diş hekimliği teşhislerinde kritik bir görevdir ve dental koşulların doğru tespiti ve analizi için büyük önem taşır. Doğru segmentasyon, diş hekimliğinde teşhis ve tedavi planlama süreçlerini önemli ölçüde iyileştirebilir. Geleneksel yöntemler genellikle doğruluk ve güvenilirlik açısından yetersiz kalmakta ve daha ileri tekniklere ihtiyaç duyulmaktadır. Derin öğrenmenin, özellikle de transformer modellerinin ortaya çıkışı, görüntü işleme görevlerinde önemli bir gelişme sunmaktadır. Bu proje, diş görüntülerinin analizinde doğruluk ve güvenilirliği artırmayı amaçlayarak, diş segmentasyonu gerçekleştirmek için Vision Transformer (ViT) modelini kullanmaktadır. Vision Transformer (ViT) modeli, uzun menzilli bağımlılıkları yakalamak için transformer mimarisini kullanan ve görüntü işleme görevlerinde üstün performans gösteren yenilikçi bir yaklaşımdır. Bu model, girdi görüntüyü sabit boyutlu yama (patch) dizilerine bölerek her bir yama üzerinde dikkat (attention) mekanizmaları uygular. Girdi görüntünün her bir parçası için bir gömülü vektör hesaplanır ve bu vektörler, pozisyon bilgileri ile birlikte transformer katmanlarına beslenir. Bu katmanlarda, çok başlı kendi kendine dikkat (multi-head self-attention) ve ileri beslemeli ağ (feed-forward network) yapılarını içeren katmanlar bulunmaktadır. Transformer katmanlarının çıktısı, iki boyutlu bir özellik haritasına yeniden düzenlenir ve ardından konvolüsyonel katmanlar ile işlenir. Son olarak, segmentasyon başlığı, diş segmentasyon haritasını üretir. Bu mimari, hem küresel bağlamı hem de yerel detayları yakalayarak doğru ve etkili diş segmentasyonu sağlar. Aşağıdaki diagram, ViT modelinin bu mimari yapısını ve veri akışını görsel olarak sunmaktadır.

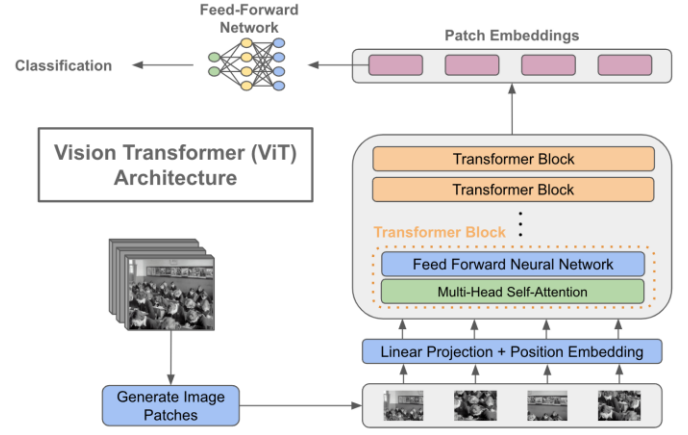


Diagram 3. ViT modelinin mimari yapısı

Vision Transformer (ViT) modeli, görüntüleri işlemek için bir dizi matematiksel işlem ve dönüşüm kullanır. İşte ViT modelinde yer alan temel formüller ve adımlar:

Yama Gömme (Patch Embedding):

- Girdi görüntü xxx boyutunda (H,W,C) olan ve her biri (P,P,C) boyutunda N yamaya bölünür.
- Her yama düzleştirilir ve D boyutlu bir alana doğrusal olarak gömülür.
- Doğrusal gömme şu şekilde ifade edilir:

$$z_0 = [x_p^1 E; x_p^2 E; \dots; x_p^N E] + E_{pos}$$

Transformer Kodlayıcı (Transformer Encoder):

- Gömülen yamalar z_0 , transformer kodlayıcıya beslenir ve bu kodlayıcı L katmandan oluşur. Her katman, çok başlı kendi kendine dikkat (multi-head self-attention) ve ileri beslemeli ağ (feed-forward network) yapılarını içerir.
- Her bir başlık hhh için kendi kendine dikkat mekanizması şu şekilde hesaplanır:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)$$

$$\text{where } d_k = \frac{d_{\text{model}}}{h}$$

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_i, \dots, \text{head}_h)W^O$$

$$\text{where } \text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V)$$

İleri beslemeli ağ şu şekilde uygulanır:

$$\text{FFN}(x) = \text{GELU}(xW_1 + b_1)W_2 + b_2$$

Her transformer katmanı şu şekilde özetlenir:

$$z' = \text{LayerNorm}(z + \text{MultiHead}(Q, K, V))$$

$$z_{l+1} = \text{LayerNorm}(z' + \text{FFN}(z'))$$

Sınıflandırma Tokeni:

- Sınıflandırma tokeni, transformer kodlayıcıya beslenmeden önce gömülen yamaların dizisine eklenir.
- Bu token, tüm yamalardan bilgi toplamak ve nihai sınıflandırma çıktısını üretmek için kullanılır.

Yeniden Şekillendirme ve Konvolüsyonel Segmentasyon Başlığı:

- Transformer kodlayıcıdan çıkan sonuç, 2D özellik haritasına yeniden şekillendirilir.
- Konvolüsyonel katmanlar, özellik haritasını rafine etmek ve nihai segmentasyon maskesini üretmek için uygulanır.
- Konvolüsyon işlemi şu şekilde verilir:

$$y = \text{Conv2D}(x, W, b)$$

Geliştirme

Bu projenin geliştirme aşaması, sağlam bir diş segmentasyon modeli oluşturmak için gerekli olan birçok kritik adımı içermektedir. Başlangıçta, 1500 panoramik ağız röntgeni görüntüsü ve bunlara karşılık gelen maskelerden oluşan bir veri seti hazırlanmıştır. Görüntüler .jpg formatında, maskeler ise dişlerin segmentasyon maskelerini temsil eden .png formatındadır. Bu bölümde, veri hazırlığı, model mimarisi tasarımı, veri artırma teknikleri, kayıp fonksiyonları, optimizasyon stratejileri ve

kapsamlı eğitim süreci ayrıntılı olarak incelenecektir.

Veri Hazırlığı

Bu projede kullanılan veri seti, görüntülerin ve maskelerin yüksek kalitede ve doğru hizalanmasını sağlamak için titizlikle hazırlanmıştır. Proje, 1500 görüntü ve bunlara karşılık gelen maskelerden oluşan bir veri seti kullanmaktadır. Görüntüler, diş kliniklerinden alınmış ve hasta gizliliğini korumak için anonimleştirilmiştir. Her görüntünün, diş sınırlarını doğru bir şekilde belirlemek için dental profesyoneller tarafından elle anotasyon yapılmış karşılık gelen bir maskesi bulunmaktadır.



Diagram 1. Orjinal görüntü

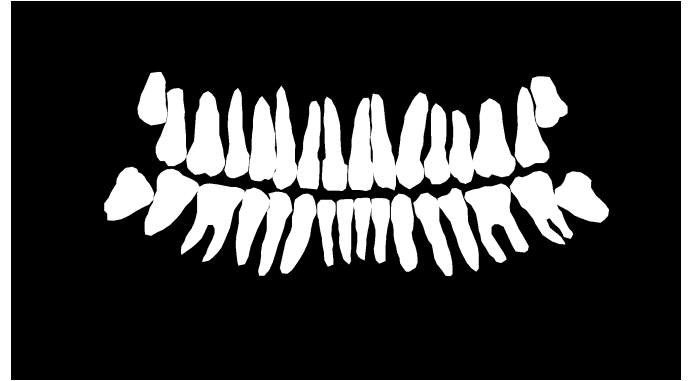


Diagram 2. Segmente edilmiş görüntü

Bu veri setini işlemek için, PyTorch'un Dataset ve DataLoader sınıfları kullanılarak özel bir veri seti sınıfı oluşturulmuştur. Bu özel sınıf, her görüntünün doğru maskesiyle eşleştirilmesini sağlamıştır. Hem görüntüler hem de maskeler, eğitim sırasında herhangi bir uyumsuzluğu önlemek için tutarlı bir sıralamada düzenlenmiştir. Bu adım, görüntüler ve maskeler arasındaki herhangi bir yanlış hizalamanın yanlış eğitime yol açabileceği ve modelin performansını olumsuz etkileyebileceği için kritik öneme sahiptir.

Görüntülerin ve maskelerin sıralanmasına ek olarak, veri ön işleme aşamasında giriş boyutunun standart hale getirilmesi sağlanmıştır. Her görüntü ve maske, Vision Transformer modelinin giriş gereksinimlerini karşılamak için 224x224 piksel boyutuna yeniden boyutlandırılmıştır. Bu yeniden boyutlandırma, modelin görüntüleri verimli bir şekilde işlemesini sağlarken, doğru segmentasyon için gerekli olan önemli özelliklerin korunmasını sağlamıştır. Aşağıdaki görsel, Vision Transformer (ViT) modelinin uygulanmasından sonraki süreci detaylandırmaktadır. Bu süreçte, konvolüsyon katmanları, havuzlama (pooling) katmanları ve

küresel ortalama (global average) katmanları gibi ek adımlar yer almaktadır. Görselde, modelin giriş görüntüsünden başlayarak ViT modelinin çıktısını nasıl işlediği ve nihai segmentasyon maskesine nasıl ulaşıldığı adım adım gösterilmektedir. Bu diyagram, ViT modelinin ardından kullanılan ek katmanlar ve işlemler hakkında daha derin bir anlayış sağlamakta ve modelin genel akışını net bir şekilde sunmaktadır:

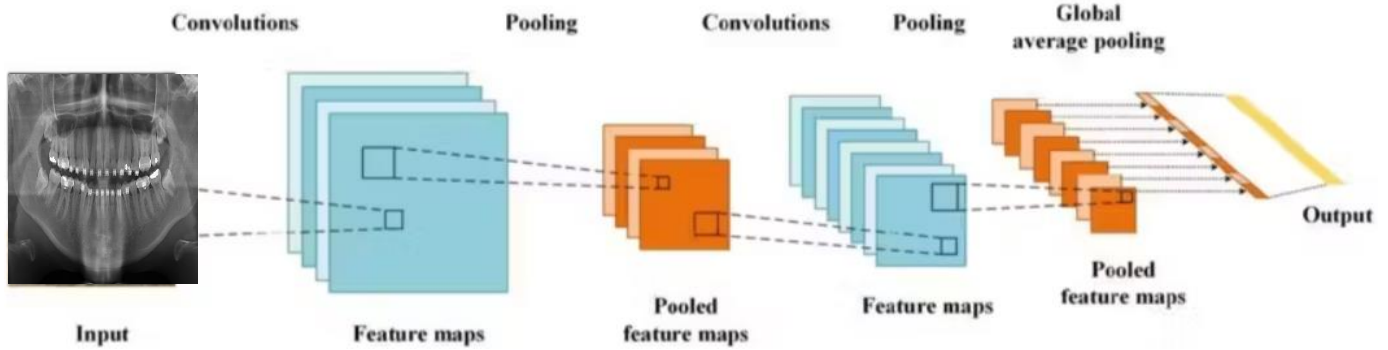


Diagram 4. Segmentasyonun genel görüntüsü

Veri Artırma

Veri artırma, eğitim verilerinin çeşitliliğini artırmada ve dolayısıyla modelin yeni, görülmemiş verilere genelleme yeteneğini iyileştirmede hayati bir rol oynamıştır. Görüntülere ve maskelere çeşitli veri artırma teknikleri uygulanmıştır. Bu teknikler arasında yatay ve dikey çevirme, rastgele döndürme, parlaklık ve kontrast ayarlamaları, gamma düzeltmeleri ve Contrast Limited Adaptive Histogram Equalization (CLAHE) bulunmaktadır.

Yatay ve dikey çevirme, modelin dişlerin çeşitli yönlerde görünebileceğini öğrenmesine yardımcı olmuştur. Rastgele döndürme, görüntü ve maskeleri rastgele açılarda döndürerek bu çeşitliliği daha da artırmış ve dişlerin farklı perspektiflerini simüle etmiştir. Parlaklık ve kontrast ayarlamaları, modelin farklı aydınlatma koşullarında çekilmiş görüntülerle başa çıkabilmesini sağlamıştır, bu da gerçek dünya senaryolarında yaygındır. Gamma düzeltmeleri, görüntülerin parlaklığını ayarlayarak modelin, parlaklık seviyeleri değişen görüntülerde dişleri ayırt etmesine yardımcı olmuştur. CLAHE, görüntülerin yerel kontrastını iyileştirerek diş sınırlarının daha belirgin hale gelmesini sağlamıştır.

Bu artırma tekniklerinin kombinasyonu, genelleme yapabilen bir model eğitmek için gerekli olan daha çeşitli bir veri seti oluşturmuştur. Modeli geniş bir dönüşüm yelpazesine maruz bırakarak, farklı koşullarda ve yönlerde dişleri tanımlamayı öğrenmiş ve bu da modelin dayanıklılığını ve doğruluğunu artırmıştır.

Model Mimarisi

Bu projenin merkezinde, görüntülerde uzun menzilli bağımlılıkları yakalama yeteneği ile bilinen Vision Transformer (ViT) modeli bulunmaktadır. ViT modeli, çeşitli görüntü işleme görevlerinde üstün performansı nedeniyle seçilmiştir. Bu projede kullanılan model, ImageNet-21k veri seti üzerinde önceden eğitilmiş olup, transfer öğrenme için sağlam bir temel sunmaktadır.

ViT model mimarisi, bir segmentasyon başlığı eklenerek daha da geliştirilmiştir. Bu segmentasyon başlığı, transformer tarafından çıkarılan özellikleri rafine eden birkaç konvolüsyon katmanından oluşmaktadır. Segmentasyon başlığı, giriş görüntüsünde dişlerin varlığını gösteren ikili bir segmentasyon haritası üretmek için tasarlanmıştır. Konvolüsyon katmanları, modelin doğru

segmentasyon için gerekli olan mekansal hiyerarşileri ve ince detayları öğrenmesine yardımcı olmuştur.

Segmentasyon başlığı, ReLU aktivasyon fonksiyonlarına sahip birden fazla konvolüsyon katmanından oluşmuştur. Bu katmanlar, özellikleri kademeli olarak rafine ederek nihai segmentasyon haritasını üretmiştir. Son katman, özellik haritasını tek bir kanala indiren ve ikili segmentasyon maskesini temsil eden 1x1 konvolüsyondur.

Bu geliştirilmiş ViT model mimarisi, karmaşıklık ve hesaplama verimliliği arasında bir denge sağlamıştır. Transformer katmanları küresel bağlamı yakalarken, konvolüsyon katmanları yerel detaylara odaklanarak, doğru dış segmentasyonu yapabilen bir model ortaya çıkmıştır.

Projede kullanılan model mimarisini aşağıdaki gibidir:

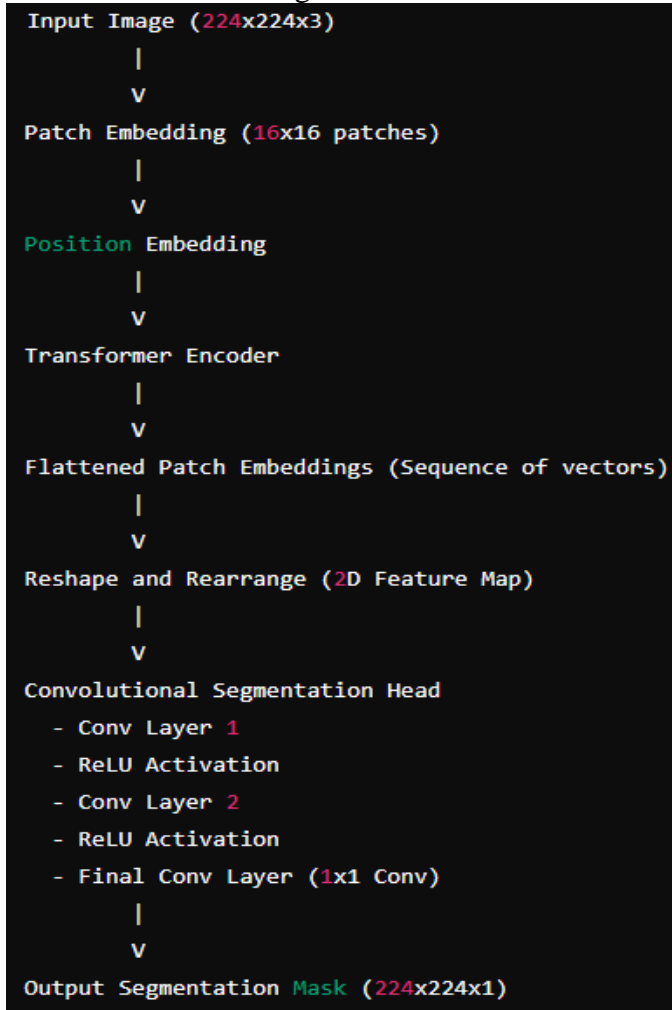


Diagram 5. Projede kullanılan model mimarisi

Kayıp Fonksiyonları

Modeli etkili bir şekilde eğitmek için birleşik bir kayıp fonksiyonu kullanılmıştır. Bu kayıp fonksiyonu, Binary Cross-Entropy (BCE) kaybı ve Dice kaybını birleştirerek her iki yöntemin güçlü yönlerinden yararlanmıştır. BCE kaybı, tahmin edilen olasılıklar ile gerçek etiketler arasındaki farkı ölçerek ikili sınıflandırma görevleri için etkilidir. Dice kaybı ise özellikle segmentasyon görevleri için kullanışlıdır, çünkü tahmin edilen ve gerçek maskeler arasındaki örtüşmeyi ölçer.

BCE kaybı, her pikselde tahmin edilen segmentasyon haritası ile gerçek maske karşılaştırılarak hesaplanmıştır. Bu kayıp fonksiyonu, yanlış tahminleri cezalandırarak modelin doğru segmentasyon haritaları üretmesini teşvik etmiştir. Ancak, BCE kaybı tek başına segmentasyon görevleri için yeterli olmayabilir, çünkü pikseller arasındaki mekansal ilişkileri hesaba katmaz.

Dice kaybı, BCE kaybını tamamlayarak tahmin edilen ve gerçek maskeler arasındaki örtüşmeye odaklanmıştır. Tahmin edilen ve gerçek maskelerin kesişimi ve birleşimi temel alınarak hesaplanmış ve iki maske arasındaki benzerliği ölçmüştür. Dice kaybını içeren model, gerçek maskelere yakın segmentasyon haritaları üretmeye teşvik edilmiştir.

Birleşik kayıp fonksiyonu, BCE kaybı ve Dice kaybının toplamı olarak tanımlanmıştır. Bu kombinasyon, modelin doğru segmentasyon haritaları üretirken pikseller arasındaki mekansal ilişkileri korumasını sağlamıştır.

Optimizasyon Stratejileri

Model ağırlıklarını güncellemek için AdamW optimizatörü kullanılmıştır. AdamW, ağırlık çürüme düzenlemesini içeren Adam optimizatörünün bir uzantısıdır. Bu düzenleme, büyük ağırlıkları cezalandırarak aşırı uyum riskini azaltır ve modelin daha basit ve genellenebilir çözümler üretmesini teşvik eder.

Optimizatör için öğrenme oranı başlangıçta 3e-5 olarak ayarlanmış ve eğitim sırasında ayarlamak için bir öğrenme oranı zamanlayıcısı kullanılmıştır. Zamanlayıcı, kayıp 5 ardışık epoch boyunca

iyileşmezse öğrenme oranını 0.5 faktörüyle azaltmıştır. Bu uyarlanabilir öğrenme oranı, modelin başlangıçta daha büyük güncellemeler yapmasına ve optimal çözüme yaklaştıkça daha ince ayarlar yapmasına olanak tanıyarak modelin daha verimli bir şekilde yakınsamasını sağlamıştır.

Optimizasyon stratejisi, yakınsama hızını ve eğitim sürecinin istikrarını dengelemek için tasarlanmıştır. AdamW ve öğrenme oranı zamanlayıcısının kombinasyonu, modelin etkili bir şekilde öğrenmesini sağlarken aşırı uyum riskini en aza indirmiştir.

Eğitim Süreci

Eğitim süreci, veri seti üzerinde 80 epoch boyunca yinelemeler yapmayı içermektedir. Her epoch sırasında, model eğitim verileri üzerinde eğitilmiş ve kayıp, Intersection over Union (IoU), Dice katsayısı ve doğruluk gibi çeşitli metrikler hesaplanmıştır. Bu metrikler, modelin performansı hakkında içgörü sağlamakta ve doğruluğu artırmak için ayarlamaları yönlendirmektedir.

Eğitim döngüsü, veri yükleme işlemini verimli bir şekilde yönetmek için PyTorch'un DataLoader sınıfını kullanarak tasarlanmıştır. Her bir batch, eğitim için modele beslenen bir dizi görüntü ve karşılık gelen maskelerden oluşmaktadır.

Eğitim sırasında, modelin tahminleri gerçek maskelerle karşılaştırılarak birleşik kayıp hesaplanmıştır. Optimizatör, bu kayba dayalı olarak model ağırlıklarını güncellemiştir. Her batch için metrikler biriktirilmiş ve tüm epoch için ortalama metrikler hesaplanmıştır. Bu metrikler, modelin performansını değerlendirmek ve gerekli ayarlamaları yapmak için kullanılmıştır.

Modeli eğitmenin yanı sıra, eğitim döngüsü ayrıca bir doğrulama adımı da içermektedir. Her epoch'tan sonra model, genel performansını ölçmek için ayrı bir doğrulama seti üzerinde değerlendirilmiştir. Doğrulama metrikleri, modelin doğruluğu hakkında tarafsız bir değerlendirme sağlamış ve aşırı uyum gibi sorunları belirlemeye yardımcı olmuştur.

Zorluklar ve İyileştirmeler

Bu modelin geliştirilmesi sırasında çeşitli zorluklarla karşılaşmıştır. Ana zorluklardan biri, görüntülerin ve maskelerin doğru eşleştirilmesini sağlamaktır. Herhangi bir yanlış hizalama, yanlış eğitime ve kötü model performansına yol açabilir. Bu sorun, görüntülerin ve maskelerin titizlikle sıralanması ve eğitimden önce hizalanmalarının doğrulanması ile ele alınmıştır.

Bir diğer zorluk, dış segmentasyonunun ince detaylarını işlemektir. Dış sınırları genellikle karmaşıktır ve doğru segmentasyon için yüksek hassasiyet gerektirir. Veri artırma teknikleri ve birleşik kayıp fonksiyonu, bu zorluğun üstesinden gelmekte kritik öneme sahiptir. Modeli geniş bir dönüşüm yelpazesine maruz bırakarak ve tahmin edilen ve gerçek maskeler arasındaki örtüşmeye odaklanarak model, doğru segmentasyon haritaları üretmeyi öğrenmiştir.

Modelin performansını daha da iyileştirmek için ek teknikler, örneğin post-processing teknikleri, keşfedilebilir. Conditional Random Fields (CRFs) gibi post-processing teknikleri, segmentasyon sınırlarını rafine etmeye ve genel doğruluğu artırmaya yardımcı olabilir. Farklı model mimarilerini ve hiperparametreleri denemek de daha iyi sonuçlar verebilir.

Sonuçlar

Bu projeden elde edilen sonuçlar, Vision Transformer modelinin dış segmentasyonu için etkinliğini göstermektedir. Model, yüksek doğruluk ve güvenilirlik gösteren umut verici metriklere ulaşmıştır. Ortalama kayıp, IoU, Dice katsayısı ve doğruluk, epoch'lar boyunca sürekli olarak iyileşmiş ve birleşik kayıp fonksiyonu ve veri artırma tekniklerinin uygulanmasından sonra belirgin iyileşmeler gözlenmiştir.

Aşağıdaki grafikler modelin performansının her epoch ile nasıl değiştiğini göstermektedir.

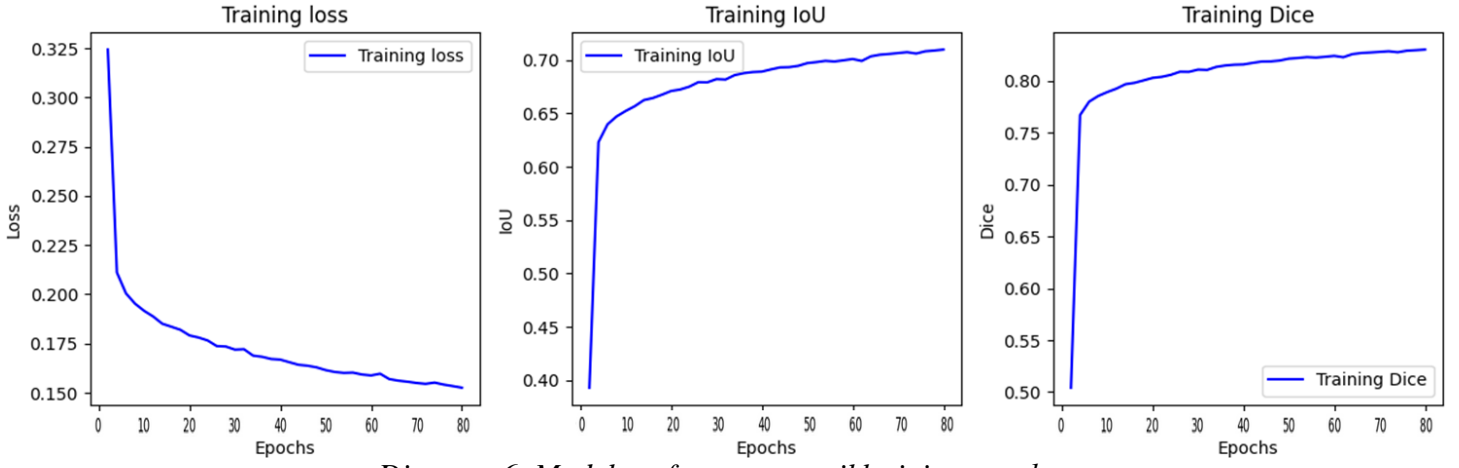


Diagram 6. Model performans metriklerinin sonuçları

Özellikle 80 epoch'tan sonra elde edilen nihai sonuçlar, modelin dişleri doğru bir şekilde segmentleme yeteneğinde önemli bir gelişme göstermiştir. Nihai metrikler şu şekildedir: Loss = 0.08, IoU = 0.722, Dice katsayısı = 0.881 ve doğruluk = 0.948. Bu sonuçlar, modelin yüksek doğrulukla dişleri çevre dokudan ayırt edebildiğini göstermektedir, ancak dişler arasındaki sınırların rafine edilmesi gibi geliştirme alanları hala mevcuttur. Segmentasyon sonuçlarının görsel incelemesi, niceliksel metrikleri daha da doğrulamıştır. Segment edilen görüntüler, çoğu durumda dişlerin belirgin sınırlarını göstermiş, ancak düşük kontrastlı alanlarda bazı zorluklar yaşanmıştır.

Genel olarak, modeli 80 epoch boyunca eğittim, ancak istediğim sonuçları alamadım. Eğitimin sonunda modelin doğruluğu ve segmentasyon yetenekleri beklediğim seviyede değildi. Bu durum, modelin daha fazla epoch boyunca eğitilmesi gerektiğini ve daha güçlü bir hesaplama altyapısının kullanılmasının performansı önemli ölçüde artırabileceğini göstermektedir. Bir sonraki aşamada, daha güçlü bir makinede 200 epoch boyunca eğiterek modelin performansını yeniden değerlendireceğim. Bu ek eğitim süresi ve gelişmiş donanım, modelin daha iyi öğrenmesini ve daha yüksek doğrulukta diş segmentasyonu yapabilmesini sağlayacaktır. Bu süreçte, modelin kayıp fonksiyonu, doğruluk, IoU ve Dice katsayısı gibi metrikler üzerinden performansı yakından takip edilerek, gerekli ayarlamalar yapılacaktır. Daha uzun bir eğitim süresi, modelin daha karmaşık desenleri öğrenmesine ve genel performansını artırmaya olanak tanıyacaktır.

Proje sırasında, özellikle görüntülerin ve maskelerin doğru eşleştirilmesini sağlama ve diş

segmentasyonunun ince detaylarını işleme konusunda zorluklarla karşılaşmıştır. Veri artırma teknikleri ve birleşik kayıp fonksiyonu, bu zorlukların üstesinden gelmekte kritik öneme sahiptir. Modelin genel performansını iyileştirmek için ön işleme adımları titizlikle tasarlanmış ve giriş görüntülerinin yüksek kalitede olmasını ve maskelerin karşılık gelen görüntülerle doğru şekilde hizalanmasını sağlamıştır. Bütün bu adımlar geçtikten sonra, model yeni bir görüntü üzerinde test edilmiştir, ve sonuç aşağıdaki gibidir.

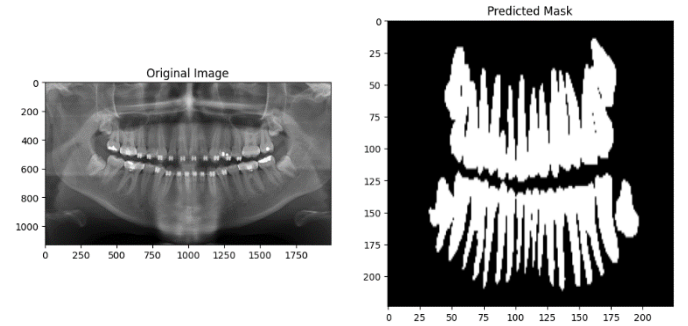


Diagram 7. Modelin yeni bir görüntü üzerinde test sonucu

Sonuç olarak, konvolüsyonel katmanlar ile geliştirilen ve birleşik kayıp fonksiyonu kullanılarak eğitilen Vision Transformer modeli, diş segmentasyonu görevlerinde önemli bir potansiyel göstermiştir. Bu projede uygulanan teknikler ve metodolojiler, diş hekimliği teşhislerinde daha fazla iyileştirme ve uygulama için sağlam bir temel oluşturmaktadır. Gelecekteki çalışmalar, model mimarisini rafine etmeye, farklı eğitim yapılandırmalarını denemeye ve daha büyük veri setlerini kullanarak modelin performansını daha da artırmaya odaklanabilir. Ayrıca, Conditional Random Fields (CRFs) gibi post-processing

tekniklerinin entegrasyonu, segmentasyon sınırlarını daha da rafine ederek modelin doğruluğunu ve klinik ortamlarda kullanılabilirliğini artırabilir.

References

1. Dosovitskiy, A., et al. (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. arXiv preprint arXiv:2010.11929.
2. Vaswani, A., et al. (2017). Attention is All You Need. Advances in Neural Information Processing Systems, 30.
3. Carion, N., et al. (2020). End-to-End Object Detection with Transformers. European Conference on Computer Vision (ECCV), 213-229.
4. Chen, L.C., et al. (2018). Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. European Conference on Computer Vision (ECCV), 801-818.
5. Kingma, D.P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. arXiv preprint arXiv:1412.6980.
6. Kolesnikov, A., et al. (2020). Big Transfer (BiT): General Visual Representation Learning. arXiv preprint arXiv:1912.11370.
7. Selvaraju, R.R., et al. (2017). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 618-626.
8. He, K., et al. (2016). Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770-778.