



**FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
YAZILIM MÜHENDİSLİĞİ**

Yüksek Lisans Tez Sunumu: Ahmet MİRA

Danışman: Doç. Dr. Fatih ÖZYURT

Uydu Görüntülerinde Görüş Transformatörleri Teknikleriyle Orman Yangınlarının Algılanmasının Geliştirilmesi

Enhancing Forest Fire Detection in Satellite Imagery Using Vision Transformers Techniques

Orman Yangınları: Ciddi Bir Tehdit, Erken Tespitle Mücadele

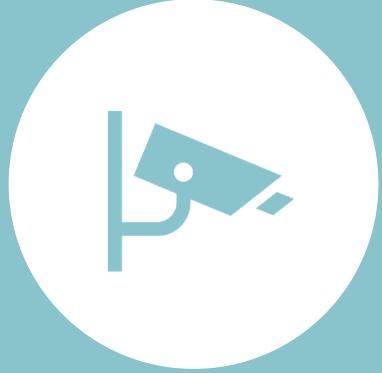
Orman yangınları, dünya genelinde ekosistemlere, biyolojik çeşitliliğe ve insan yaşamına ciddi zararlar veren yıkıcı doğal afetlerdir. Bu yangınlar, geniş orman alanlarını yok ederek doğal yaşam alanlarının kaybına, toprak erozyonuna ve ciddi hava kirliliğine yol açar. Ekonomik açıdan büyük kayıplara neden olmanın yanı sıra, yerleşim yerlerini tehdit ederek can ve mal kaybına da yol açabilirler.

Yangınların erken tespiti, etkili müdahale için hayatı önem taşımaktadır. Zamanında tespit, hasarı en aza indirmek ve can kaybını önlemek için kritik öneme sahiptir. Hızlı müdahale, yangınların daha geniş alanlara yayılmasını kontrol altına alma ve söndürme şansını artırarak olumsuz etkilerini azaltır.



Geleneksel Yangın Tespiti





İnsan Gözlemcileri

Yangın kulelerinde veya uçaklarda görev yapan kişiler. Ancak yorgunluk ve dikkat dağınlığına yatkınlık, gece ve olumsuz hava koşullarında sınırlı görüş gibi sınırları vardır.



Yangın Kuleleri

Stratejik konumlara yerleştirilmiş yapılar. Ancak sınırlı kapsama alanı ve coğrafi engeller görüşü engelleyebilir.



Devriyeler

Uçaklar veya kara araçlarıyla belirli bölgeleri kontrol etme. Ancak maliyetli ve zaman alıcı bir yöntem olup, olumsuz hava koşullarında etkisiz kalabilir.

Yapay Zeka ve Bilgisayar Görü ile Çözüm

Yapay zeka, karmaşık problemleri çözme,örüntülerin tanıma ve insan müdahalesi olmadan kararlar alma yeteneği ile bu süreci destekler.

Bilgisayar görüp, yapay zekanın önemli bir alt dalıdır ve bilgisayarların görsel verileri anlamayı yeteneğini geliştirmeyi amaçlar.

Orman yangınlarının tespitinde, uydu görüntülerinin otomatik analizi için kullanılır.



Makine Öğrenmesi ile Otomatik Tespit

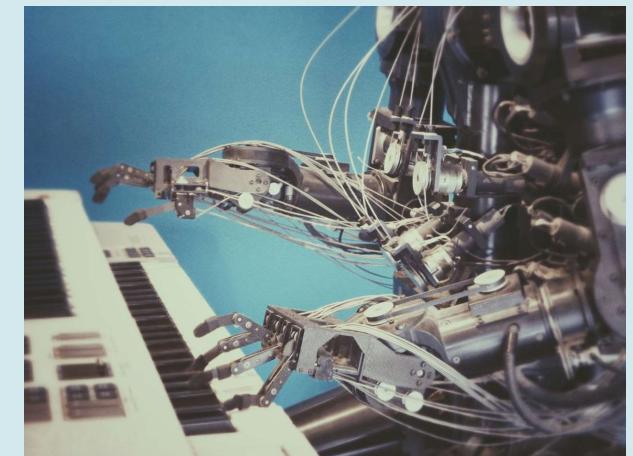
Makine Öğrenmesi (ML)

Bilgisayarların açıkça programlanmadan verilerden öğrenmesini sağlayan bir YZ alt alanı.

Gözetimli Öğrenme: Etiketli verilerle (örnek: yanın olan ve olmayan görüntüler) modeli eğitme.

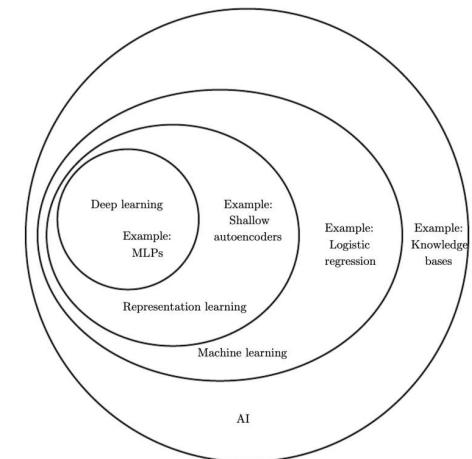
Gözetimsiz Öğrenme: Etiketsiz verilerdeki gizli kalıpları ve yapıları keşfetme (örnek: yanın yayılma biçimini analiz etme).

Pekiştirmeli Öğrenme: Modelin bir ortamda etkileşim kurarak ve geri bildirim alarak öğrenme (örnek: yanın söndürme robotları).

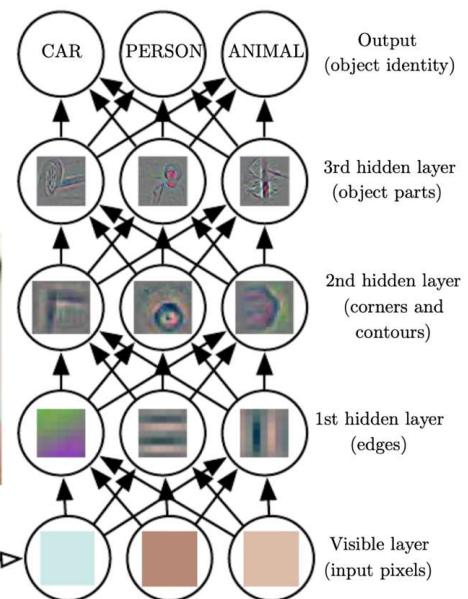


Derin Öğrenme

Derin öğrenme, karmaşık kalıpları öğrenmek için çok katmanlı sinir ağları kullanan bir makine öğrenmesi alt alanıdır. Çok katmanlı mimariler, otomatik öznitelik çıkarımı, doğrusal olmayanlık ve özel öğrenme algoritmaları gibi özelliklere sahiptir. Bu özellikler, derin öğrenme modellerinin yüksek doğruluk, esneklik ve ölçülebilirlik sağlamasına olanak tanır.



Derin Öğrenme



Özellikler

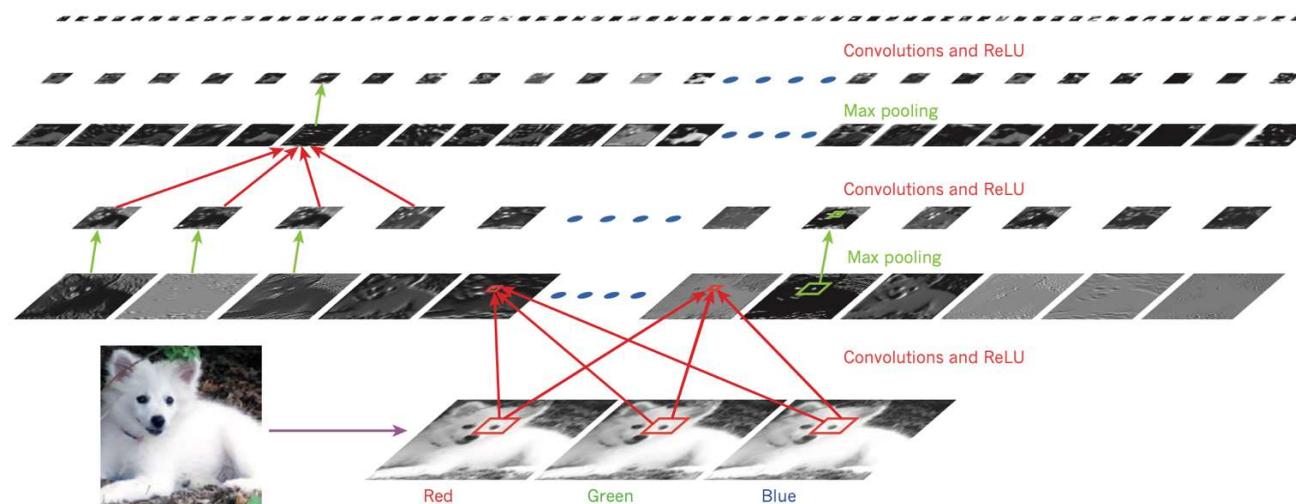
- Çok Katmanlı Mimariler
- Otomatik Öznitelik Çıkarımı
- Doğrusal Olmayanlık
- Öğrenme Algoritmaları

Avantajlar

- Yüksek doğruluk
- Esneklik
- Ölçeklenebilirlik
- Farklı veri türleriyle çalışma

Evrişimli Sinir Ağları (CNN)

Evrişimli Sinir Ağları (CNN), görüntülerdeki mekansal özellikleri çıkarmak için özelleştirilmiş bir derin öğrenme mimarisidir. CNN'ler, evrişim katmanı, aktivasyon fonksiyonu, havuzlama katmanı, tam bağlı katman ve toplu normalleştirme gibi çeşitli katmanlardan oluşur. Bu katmanlar, görüntülerdeki yerel özellikleri çıkarmak, doğrusal olmayan ilişkileri modellemek ve özellik haritalarının boyutunu küçültmek gibi işlevleri yerine getirir.



Evrişimli Sinir Ağları (CNN)



Evrişim Katmanı

Yerel özellikleri çıkarmak için filtreler kullanır



Aktivasyon Fonksiyonu

Doğrusal olmayan ilişkileri modellemeye olanak tanır



Havuzlama Katmanı

Özellik haritalarının boyutunu küçültür



Tam Bağlı Katman

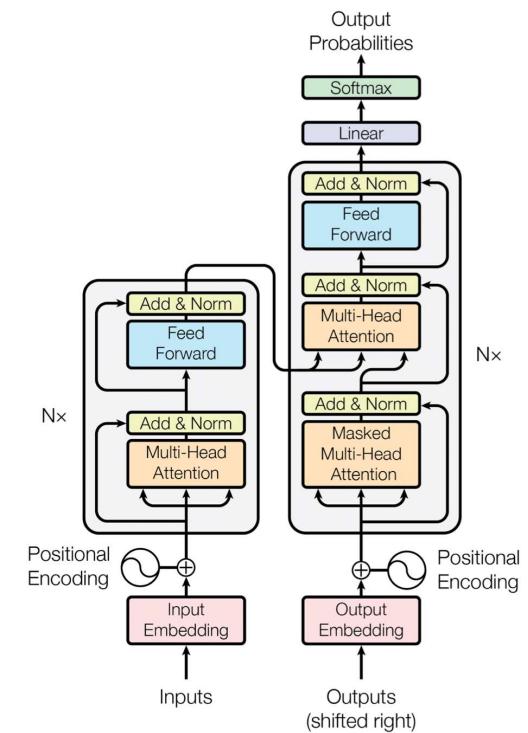
Sınıflandırma veya regresyon için kullanılır

Doğal Dil İşleme (NLP)

Doğal Dil İşleme (NLP), bilgisayarların insan dilini anlaması, yorumlaması ve üretmesini sağlayan bir yapay zeka alanıdır. NLP, insanların günlük yaşamlarında kullandığı doğal dilleri (örneğin, İngilizce, Türkçe, Çince) analiz etmek, anlamak ve üretmek için algoritmalar ve teknikler geliştirir. NLP, metin analizi, duygusal analizi, makine çevirisisi, konuşma tanıma ve soru cevaplama gibi çeşitli uygulamalarda kullanılır. Son yıllarda, derin öğrenme modellerinin gücüyle NLP alanında büyük ilerlemeler kaydedilmiştir. NLP, büyük miktarda metin verisinden anlamlı bilgiler çıkarabilir ve karmaşık dil görevlerini yerine getirebilir.

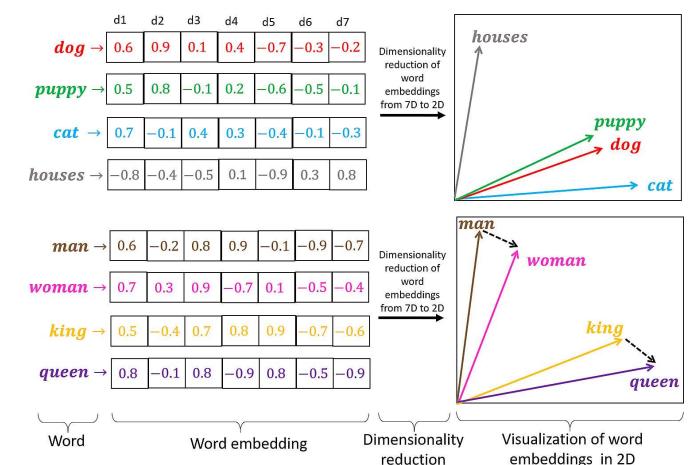
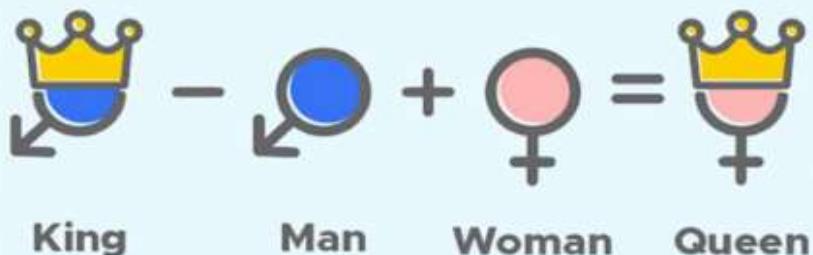
Transformatör (Transformer)

Transformatör, NLP için geliştirilmiş bir derin öğrenme mimarisidir ve RNN'lerin uzun mesafeli bağımlılıkları öğrenmedeki sınırlamalarını aşar. Transformatör, dikkat mekanizması (self-attention) kullanarak girdi dizisindeki tüm kelimeler arasındaki ilişkileri doğrudan modelleyebilir. Bu sayede, transformatör modelleri, RNN'lerden daha hızlı ve doğru sonuçlar elde edebilir.

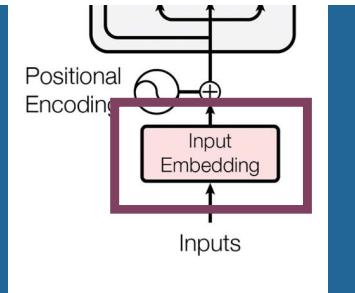


Gömmeler (Embeddings)

Gömmeler, kelimeleri veya diğer ayrik verileri sürekli vektör uzayında temsil etmek için kullanılır. Gömme vektörleri, kelimelerin anlamsal ve sözdizimsel özelliklerini yakalar ve derin öğrenme modelleri tarafından işlenebilir. Örneğin, "kedi" ve "köpek" kelimeleri anlam olarak birbirine yakın olduğu için gömme vektörleri de birbirine yakın olacaktır.

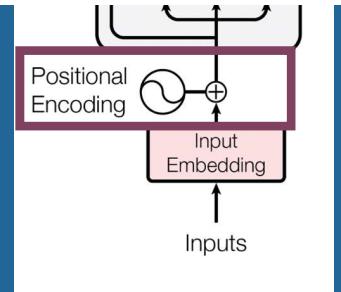


Gömmeler (Embeddings)



| Original Sentence (tokens) | YOUR | CAT | IS | A | LOVELY | CAT |
|---|---|--|---|--|--|--|
| Input IDs (Position in the Vocabulary) | 105 | 6587 | 5475 | 3578 | 65 | 6587 |
| Embedding (vector of size 512) | 952.207 5450.840 1853.448 ... 1.658 2671.529 | 171.411 3276.350 9192.819 ... 3633.421 8390.473 | 621.659 1304.051 0.565 ... 7679.805 4506.025 | 776.562 5567.288 58.942 ... 2716.194 5119.949 | 6422.693 6315.080 9358.778 ... 2141.081 735.147 | 171.411 3276.350 9192.819 ... 3633.421 8390.473 |

Konumsal Kodlama (Positional Encoding)



Transformatör modelleri, sıralı verilerle çalışırken kelimelerin dizideki konum bilgisini de kullanır. Konumsal kodlama, kelimelerin konum bilgilerini temsil eden vektörler oluşturarak modele sunar. Bu sayede, model, kelimelerin dizideki sırasını dikkate alarak daha doğru tahminler yapabilir. Örneğin, "kedi köpeği kovaladı" ve "köpek kediyi kovaladı" cümleleri farklı anlamlara sahip olduğu için, konumsal kodlama modelin bu farkı anlamasını sağlar.

YOUR CAT IS

$$PE(pos, 2i) = \sin \frac{pos}{10000^{\frac{2i}{d_{model}}}}$$

$$PE(pos, 2i + 1) = \cos \frac{pos}{10000^{\frac{2i}{d_{model}}}}$$

| | |
|-----------|-----------|
| PE(0,0) | PE(1,0) |
| PE(0,1) | PE(1,1) |
| PE(0,2) | PE(1,2) |
| ... | ... |
| PE(0,510) | PE(1,510) |
| PE(0,511) | PE(1,511) |

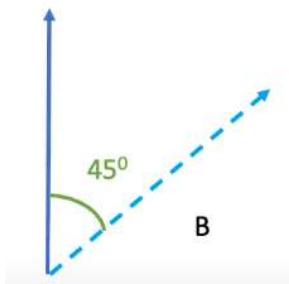
Sentence 1

Konumsal Kodlama (Positional Encoding)

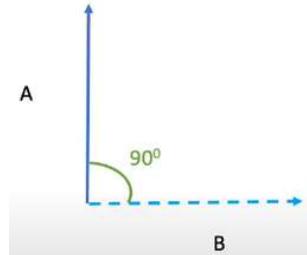
| Original Sentence | YOUR | CAT | IS | A | LOVELY | CAT |
|---|---|---|---|--|--|---|
| Embedding (vector of size 512) | 952.207 5450.840 1853.448 ... 1.658 2671.529 | 171.411 3276.350 9192.819 ... 3633.421 8390.473 | 621.659 1304.051 0.565 ... 7679.805 4506.025 | 776.562 5567.288 58.942 ... 2716.194 5119.949 | 6422.693 6315.080 9358.778 ... 2141.081 735.147 | 171.411 3276.350 9192.819 ... 3633.421 8390.473 |
| Position Embedding (vector of size 512). Only computed once and reused for every sentence during training and inference | ... | 1664.068 8080.133 2620.399 ... 9386.405 3120.159 | ... | ... | ... | 1281.458 7902.890 912.970 3821.102 1659.217 7018.620 |
| Encoder Input (vector of size 512) | ... | 1835.479 11356.483 11813.218 ... 13019.826 11510.632 | ... | ... | ... | 1452.869 11179.24 10105.789 ... 5292.638 15409.093 |

Benzerlik (Similarity)

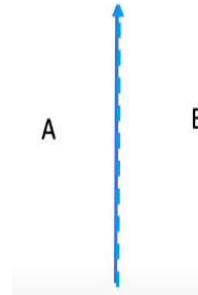
Dikkat mekanizmasının temelini oluşturan benzerlik ölçümü, iki vektör arasındaki ilişkiyi veya yakınlığı belirlemek için kullanılır. Doğal dil işlemede yaygın olarak kullanılan bir benzerlik ölçüsü kosinüs benzerliği (cosine similarity)'dır



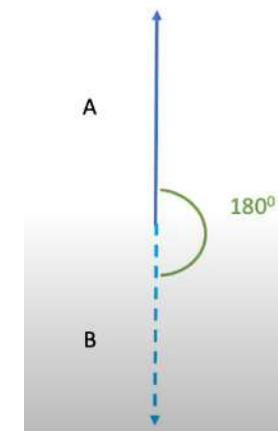
$$\text{Cos}(45) = +0.7$$



$$\text{Cos}(90) = 0$$



$$\text{Cos}(0) = +1$$



$$\text{Cos}(180) = -1$$

Benzerlik (Similarity)

Kosinüs Benzerliği

Kosinüs benzerliği, iki vektör arasındaki açının kosinüsünü hesaplayarak benzerliklerini ölçer. Açı ne kadar küçükse, vektörler o kadar benzerdir.

Kosinüs benzerliği, -1 ile 1 arasında değerler alır:

1: Vektörler tamamen aynı yönindedir (en yüksek benzerlik).

0: Vektörler ortogonaldır (ilintisiz).

-1: Vektörler tamamen zıt yönindedir (en düşük benzerlik).

$$\text{Cos } (A, B) = \frac{A \cdot B}{|A| |B|}$$

İki vektör A ve B için kosinüs benzerliği şu şekilde hesaplanır:

A · B: Vektörlerin nokta çarpımı.

||A|| ve ||B||: Vektörlerin uzunlukları (normları).

$$\text{similarity } (A, B) = \frac{A \cdot B^{\textcolor{red}{T}}}{\textit{scaling}}$$

Pay attention

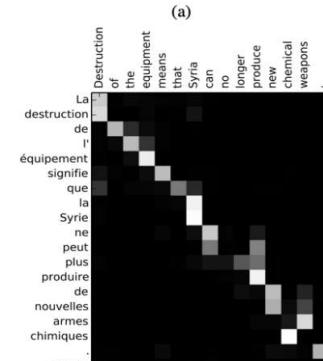
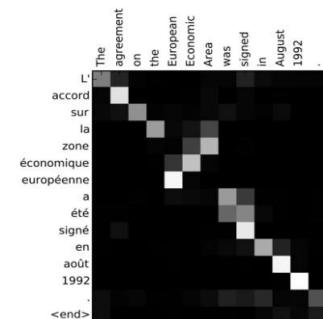
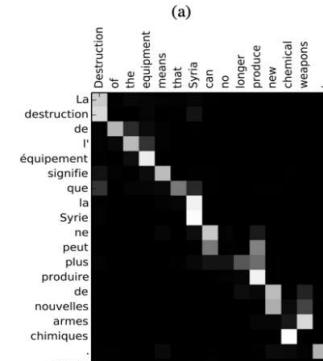
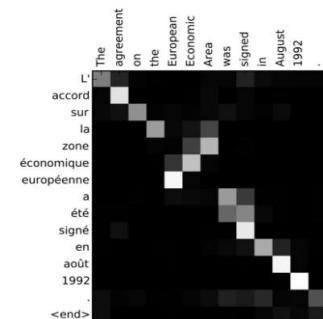
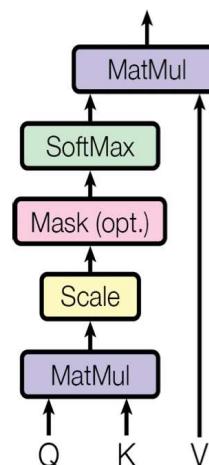


to attention.

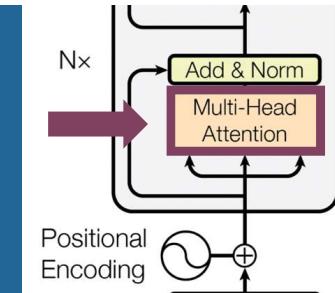
Dikkat Mekanizması

Dikkat mekanizması, transformatör modelinin temel yapı taşıdır. Bu mekanizma, girdi dizisindeki tüm kelimeler arasındaki ilişkileri hesaplar ve her bir kelimenin diğer kelimelere ne kadar "dikkat ettiğini" gösteren ağırlıklar üretir. Dikkat mekanizması, modelin uzun mesafeli bağımlılıkları öğrenmesini ve daha anlamlı temsiller oluşturmasını sağlar. Örneğin, "kedi sütü içti" cümlesinde, "içti" fiili "süt" nesnesine daha fazla dikkat eder.

Scaled Dot-Product Attention



Self-Attention: (Q,K,V)



Self-attention, bir dizideki her bir kelimenin diğer tüm kelimelerle olan ilişkisini hesaplar ve her kelime için bağılamsal bir temsil oluşturur. Bu işlem, aşağıdaki adımlarla gerçekleştirilir:

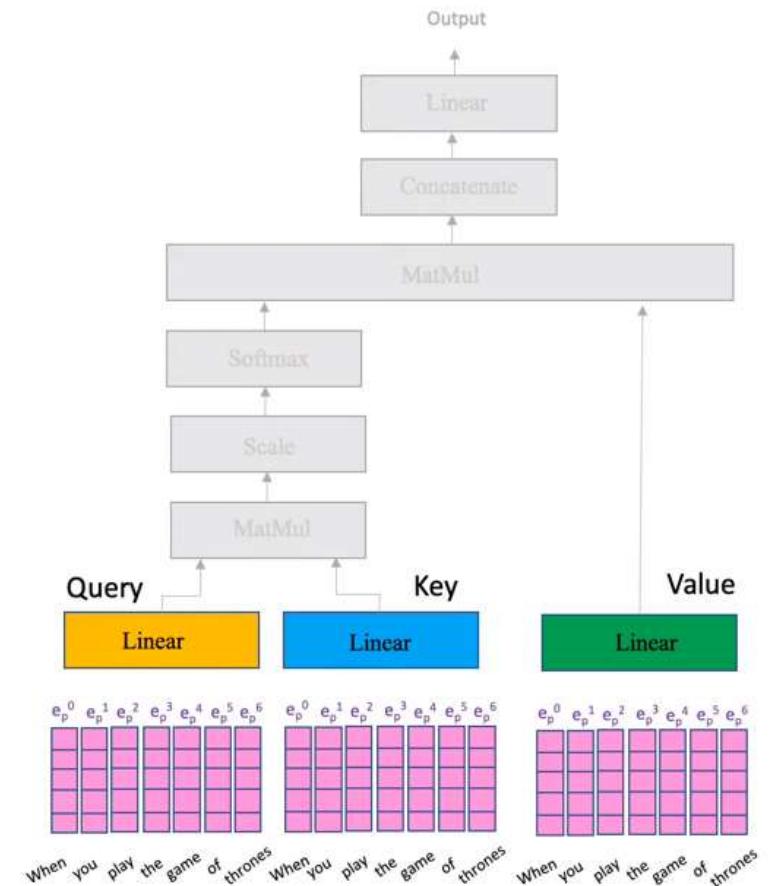
Query (Q), Key (K) ve Value (V) Vektörlerinin Oluşturulması:

Her kelime için, girdi gömülmesinden (embedding) doğrusal dönüşümler uygulanarak Q, K ve V vektörleri oluşturulur.

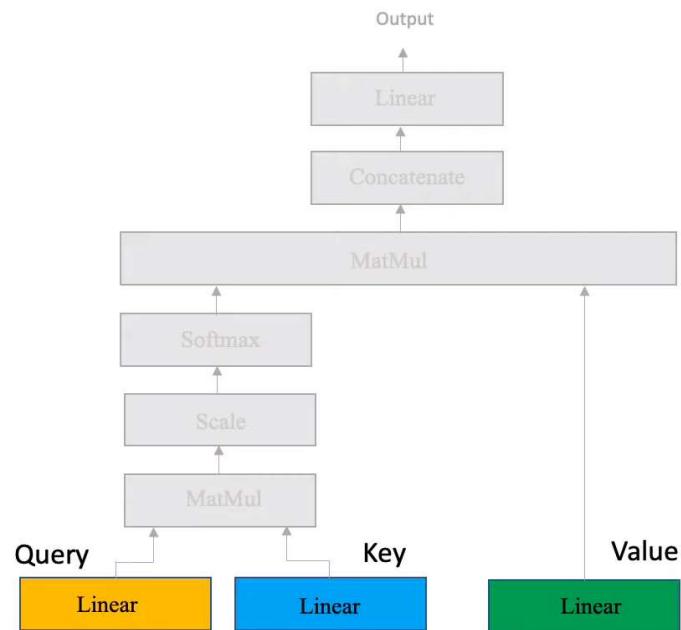
Q vektörü, kelimenin "ne aradığını" temsil eder.

K vektörü, kelimenin diğer kelimelerle nasıl ilişkili olduğunu temsil eder.

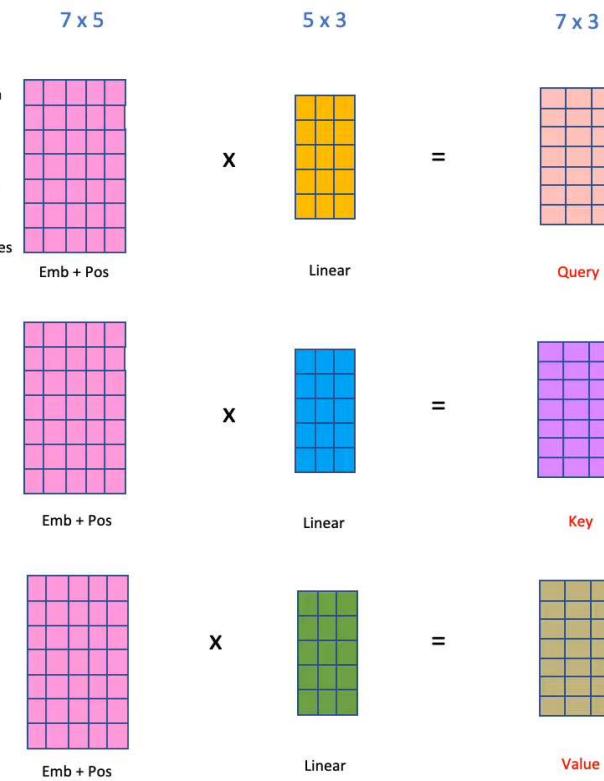
V vektörü, kelimenin içeriği bilgisi temsil eder.



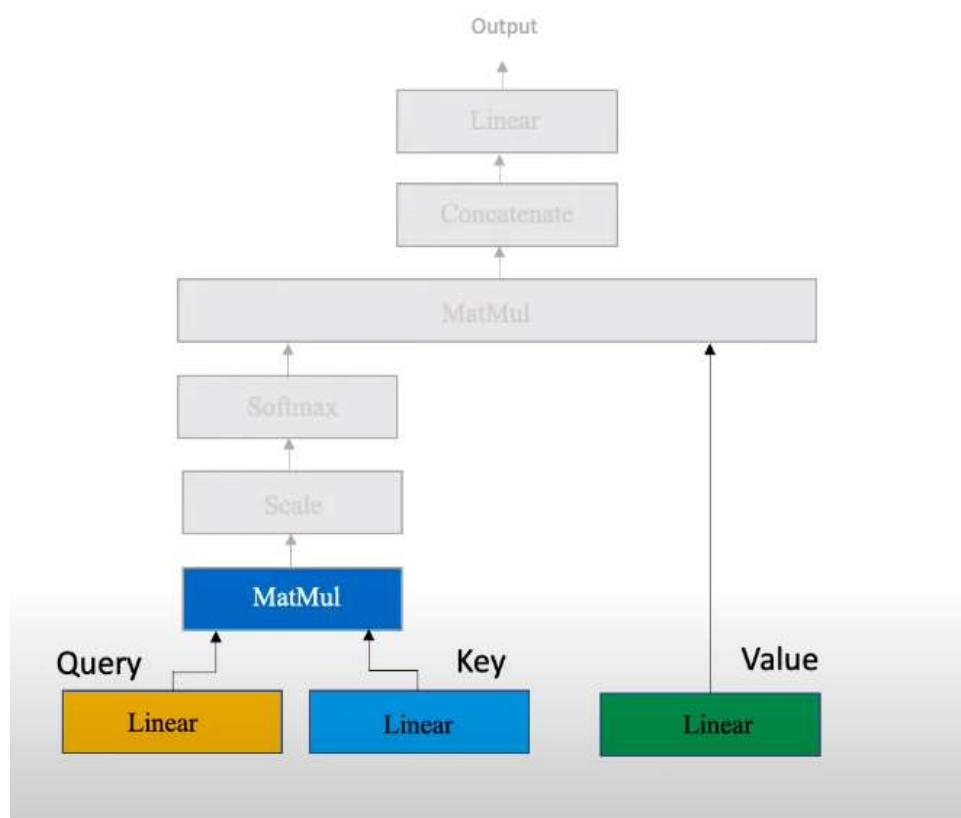
Self-Attention: Linear Katmanı



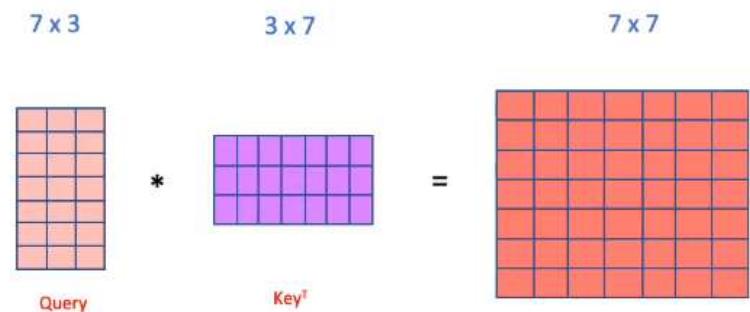
Multi-Head Attention



Self-Attention: MatMul Fonksiyonu



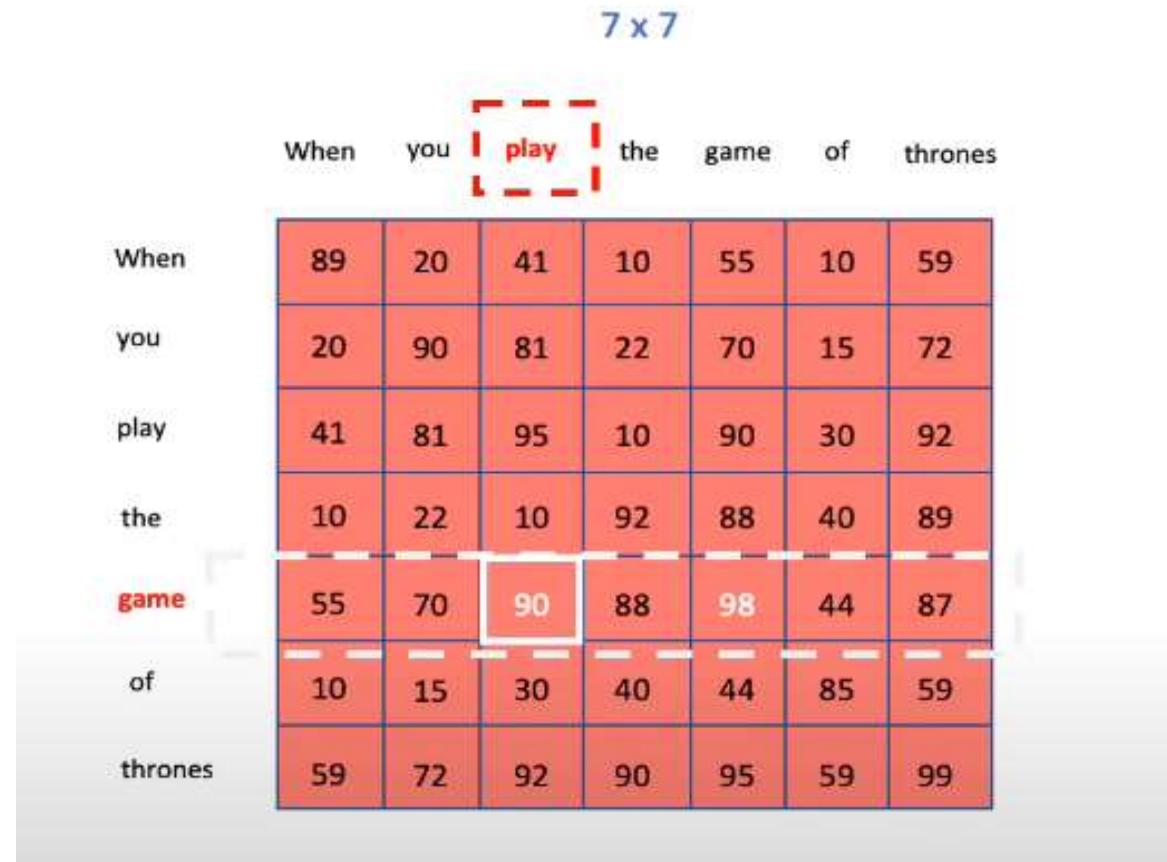
Multi-Head Attention



Attention Filter

$$\text{similarity } (Q, K) = \frac{Q \cdot K^T}{\text{scaling}}$$

Self-Attention: Attention Filter



Self-Attention: Scale Fonkisiyonu

7×7

| | When | you | play | the | game | of | thrones |
|---------|------|-----|------|-----|------|----|---------|
| When | 89 | 20 | 41 | 10 | 55 | 10 | 59 |
| you | 20 | 90 | 81 | 22 | 70 | 15 | 72 |
| play | 41 | 81 | 95 | 10 | 90 | 30 | 92 |
| the | 10 | 22 | 10 | 92 | 88 | 40 | 89 |
| game | 55 | 70 | 90 | 88 | 98 | 44 | 87 |
| of | 10 | 15 | 30 | 40 | 44 | 85 | 59 |
| thrones | 59 | 72 | 92 | 90 | 95 | 59 | 99 |

$$\text{similarity } (Q, K) = \frac{Q \cdot K^T}{\text{scaling}}$$

7×7

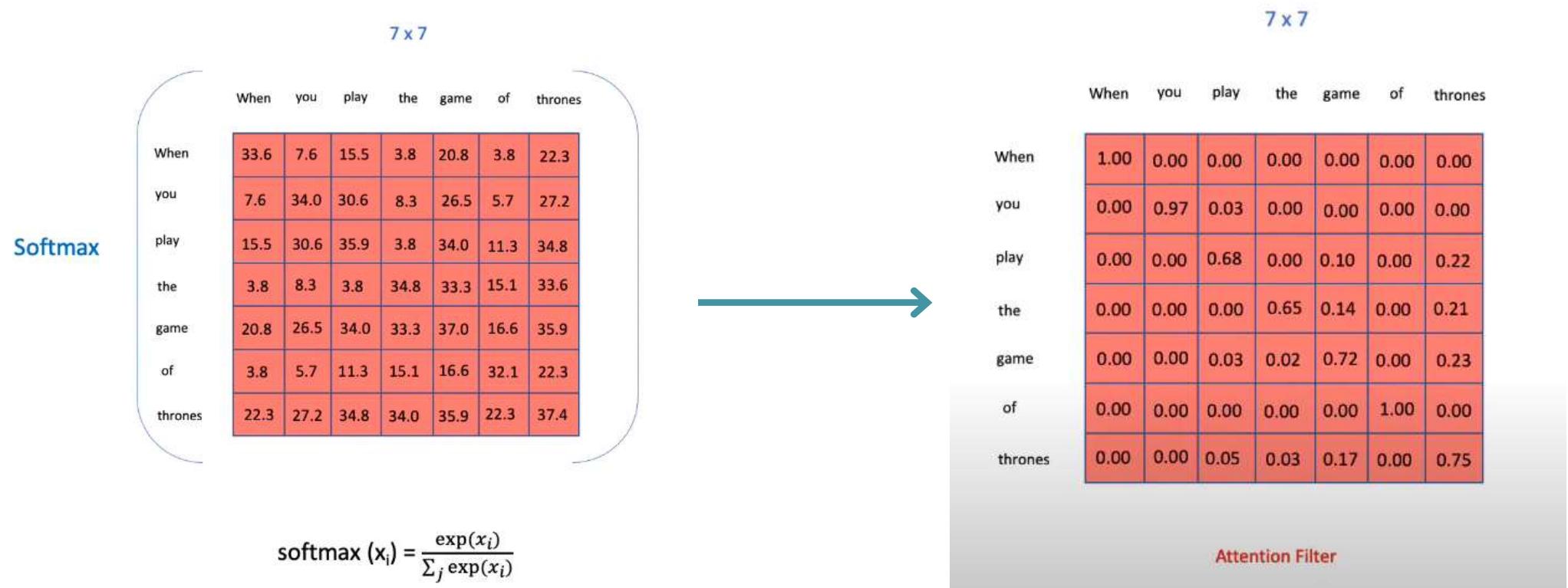
| | When | you | play | the | game | of | thrones |
|---------|------|-----|------|-----|------|----|---------|
| When | 89 | 20 | 41 | 10 | 55 | 10 | 59 |
| you | 20 | 90 | 81 | 22 | 70 | 15 | 72 |
| play | 41 | 81 | 95 | 10 | 90 | 30 | 92 |
| the | 10 | 22 | 10 | 92 | 88 | 40 | 89 |
| game | 55 | 70 | 90 | 88 | 98 | 44 | 87 |
| of | 10 | 15 | 30 | 40 | 44 | 85 | 59 |
| thrones | 59 | 72 | 92 | 90 | 95 | 59 | 99 |



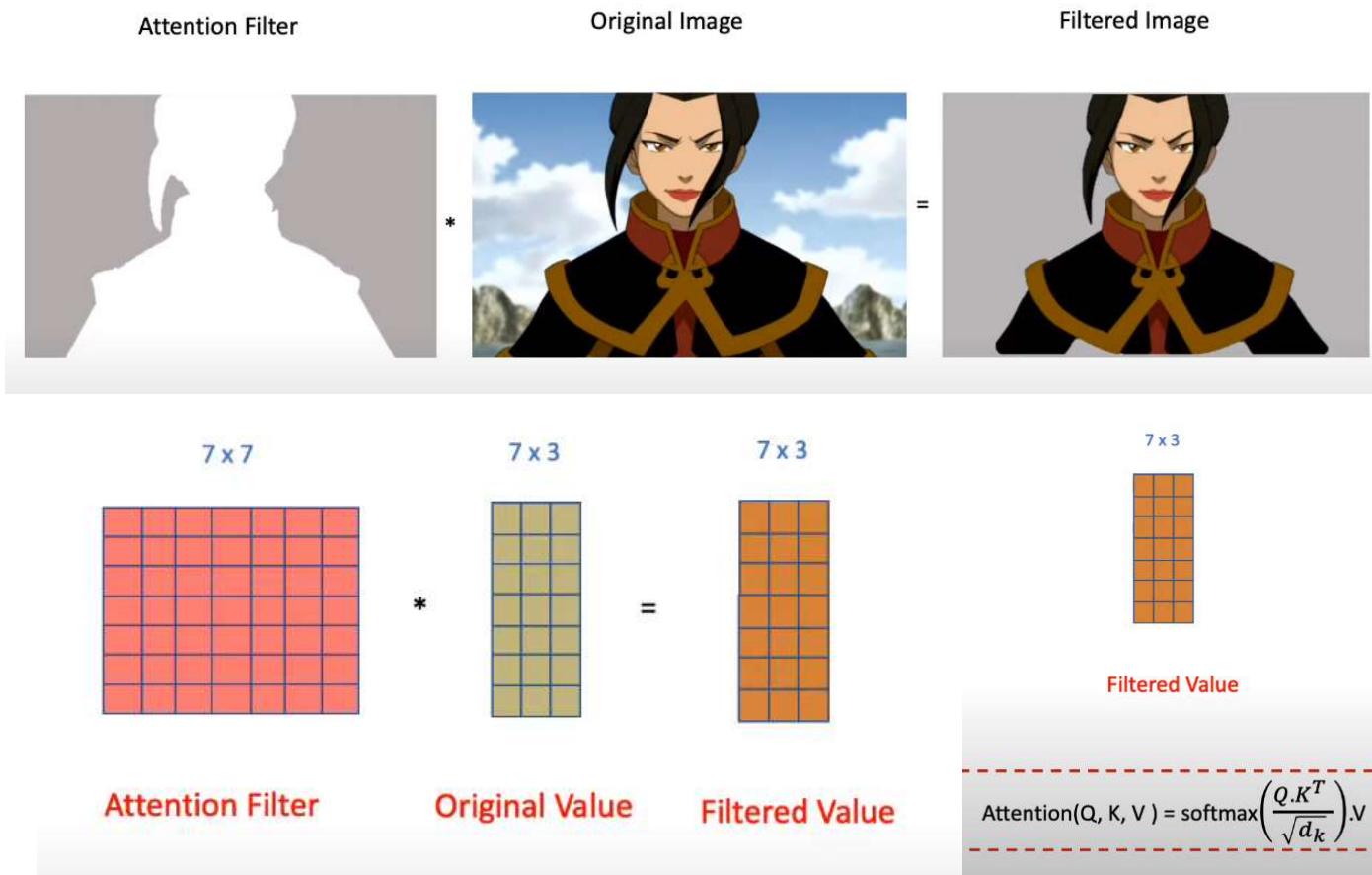
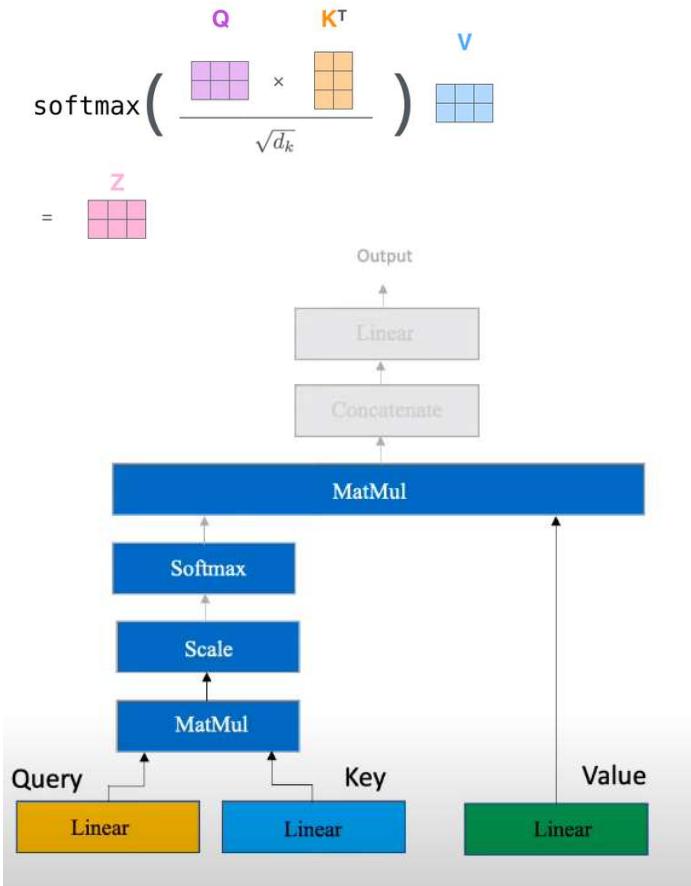
$$\sqrt{d_k}$$

$$\sqrt{7}$$

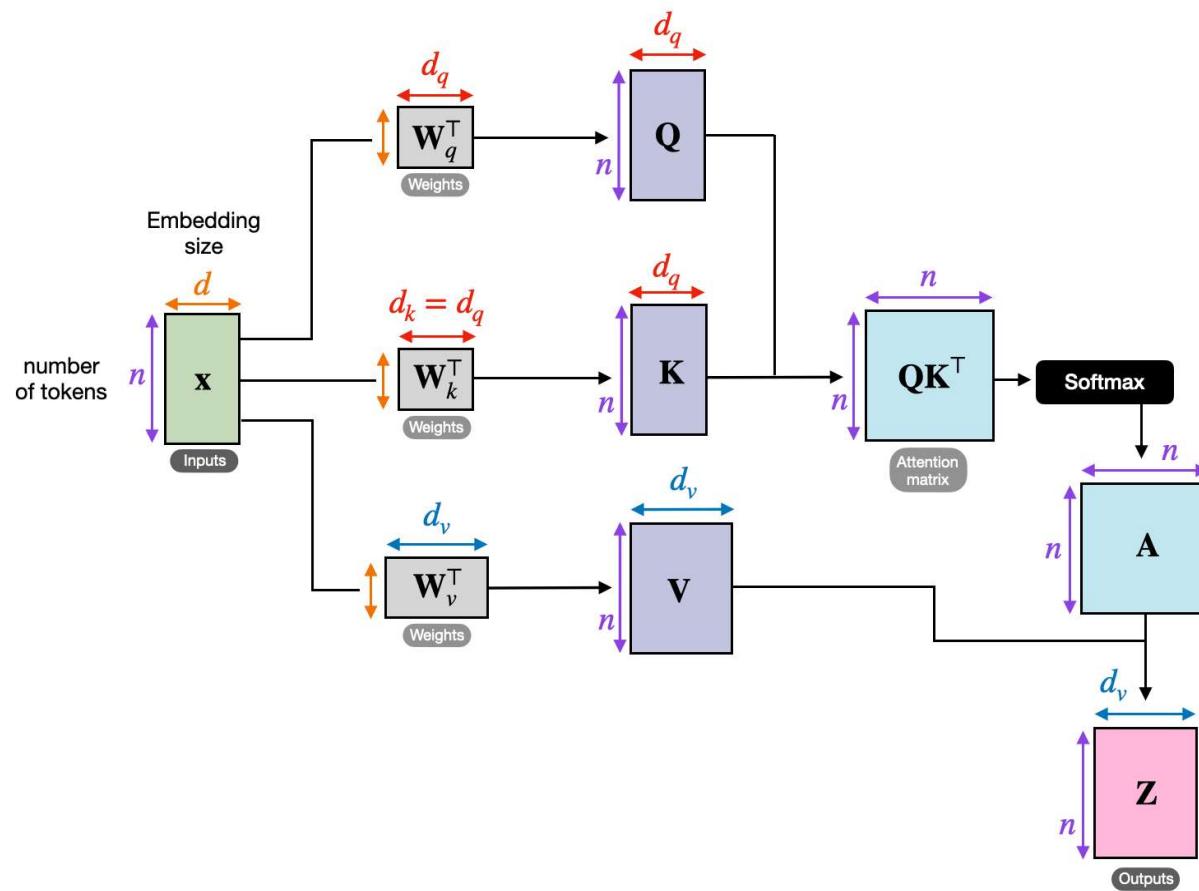
Self-Attention: SoftMax Fonkisyonu



Self-Attention

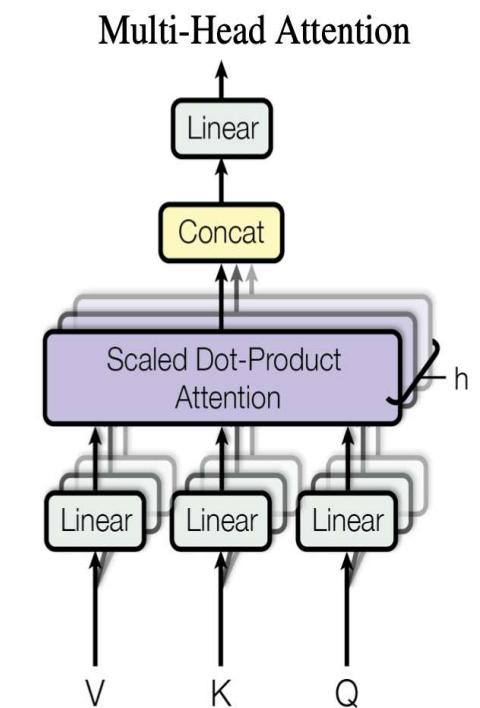


Self-Attention

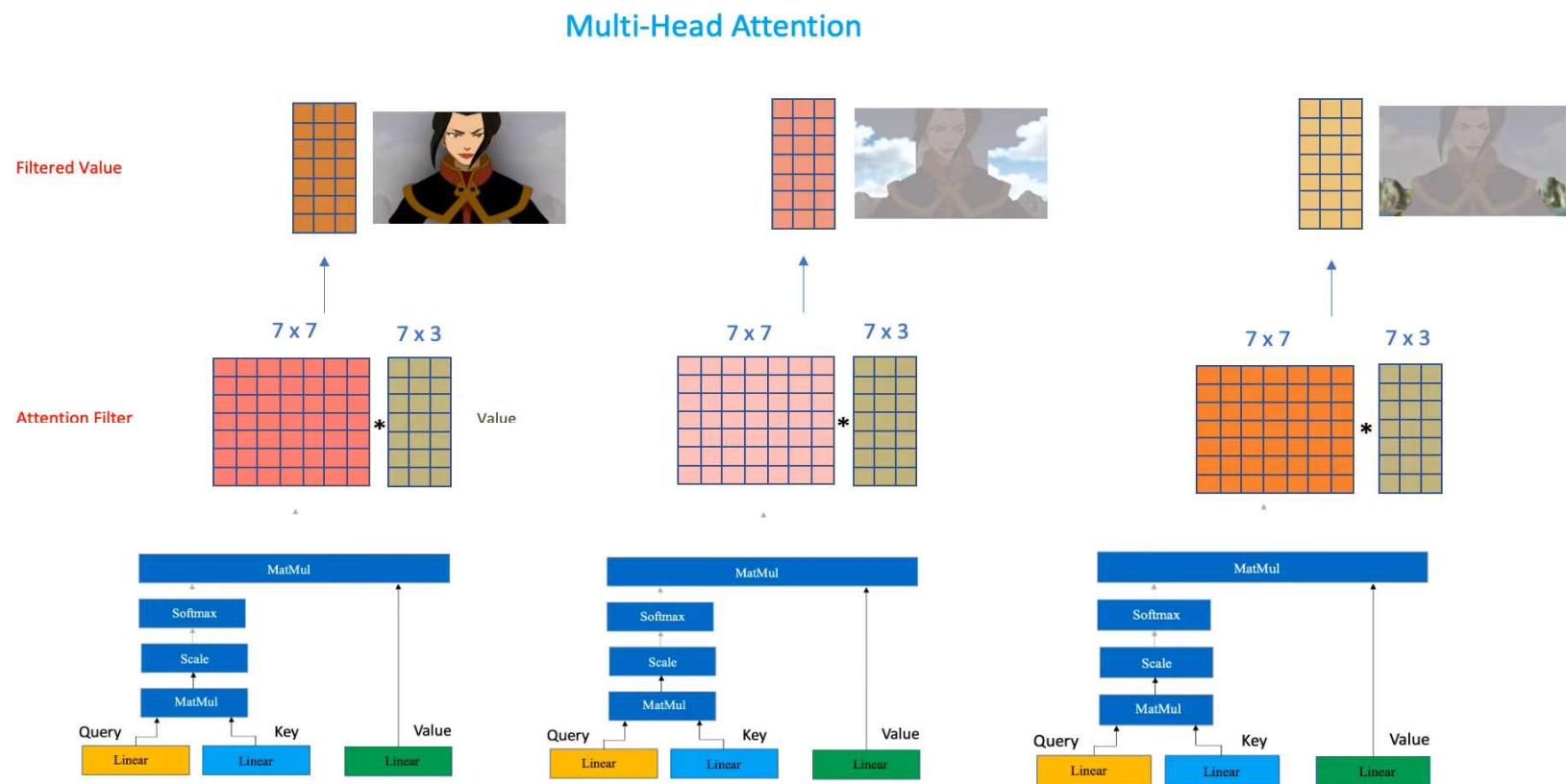


Multi-Head Attention

Multi-head attention, self-attention mekanizmasının birden fazla "kafa" ile paralel olarak uygulanmasıdır. Her kafa, farklı Q, K ve V matrisleri kullanarak farklı bir alt uzayda dikkat mekanizmasını çalıştırır. Bu sayede, model farklı yönlerden kelimeler arasındaki ilişkiyi yakalayabilir.

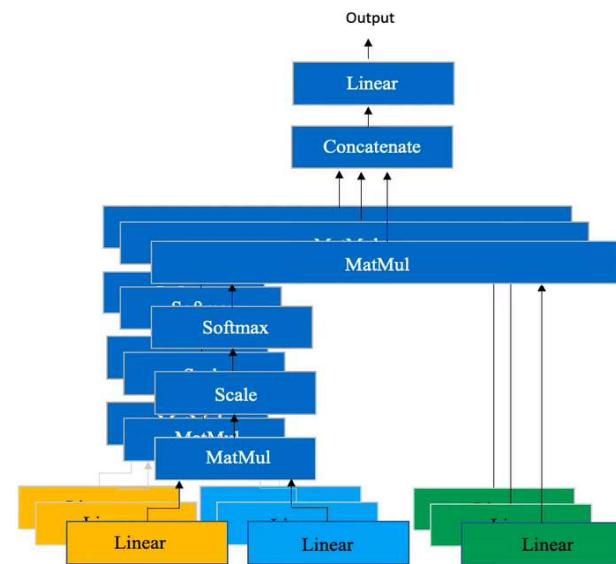


Multi-Head Attention



Multi-Head Attention

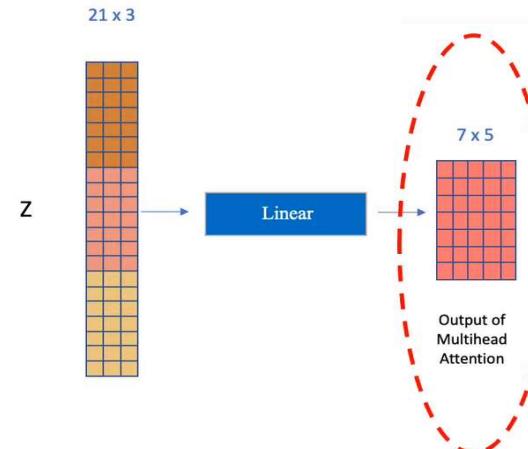
Farklı Dikkat Kafaları: Her kafa için, farklı doğrusal dönüşümler uygulanarak Q, K ve V matrisleri oluşturulur.



Multi-Head Attention

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

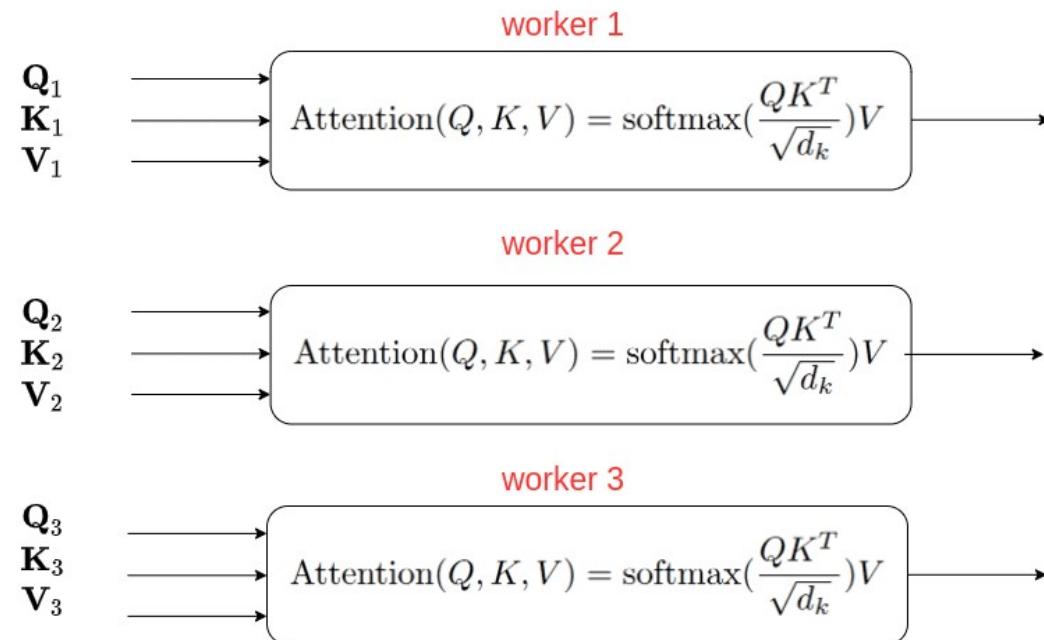
$$\begin{aligned} \text{MultiHead}(Q, K, V) &= \text{Concat}(\text{head}_1 \dots \text{head}_h)W^O \\ \text{head}_i &= \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \end{aligned}$$



Multi-Head Attention

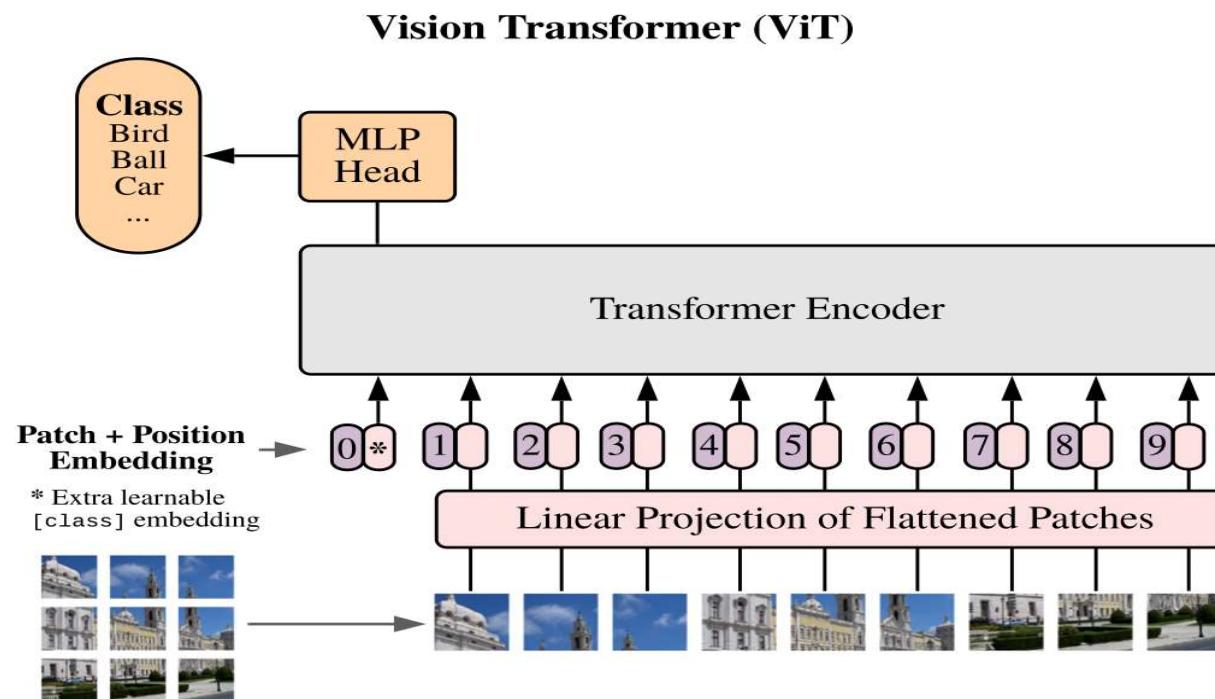
Paralel Self-Attention: Her kafa için, yukarıda açıklanan self-attention işlemi uygulanır ve farklı bağlamsal temsiller elde edilir.

Each attention head can be implemented in parallel



ViT Mimarisi

ViT mimarisi, aşağıdaki bileşenlerden oluşur:



ViT Mimarisi

Görüntü Parçalama (Patch Embedding)

Girdi görüntüsü, küçük parçalara ayrılır ve her bir parça, doğrusal bir dönüşümle bir vektör temsiline dönüştürülür. Bu dönüşüm genellikle bir evrişim işlemi ile gerçekleştirilir. Parçaların boyutu ve sayısı, modelin hiperparametreleri olarak ayarlanır. Parçalama işlemi, görüntünün mekansal bilgisini koruyarak öznitelik çıkarımını kolaylaştırır. Her bir parça, tek bir token olarak ele alınır ve bu tokenlar, tıpkı NLP'deki kelimeler gibi bir dizi halinde işlenir. Bu, ViT'nin görüntülerle NLP modellerine benzer şekilde çalışmasını sağlar.

ViT Mimarisi

Konumsal Kodlama (Positional Encoding)

Parçaların dizideki konum bilgisi, gömülü vektörlere eklenir. Bu sayede, model, parçaların göreceli konumlarını dikkate alarak daha doğru tahminler yapabilir. Konumsal kodlama, genellikle sinüs ve kosinüs fonksiyonları kullanılarak oluşturulan vektörlerle gerçekleştirilir. Bu vektörler, parçaların dizideki konumunu temsil eden farklı frekanslarda sinyaller içerir. Konumsal kodlama, transformatör modelinin sıralı verilerle çalışırken parçaların sırasını anlamasını sağlar, çünkü dikkat mekanizması, tek başına parçaların konum bilgilerini içermez.

ViT Mimarisi

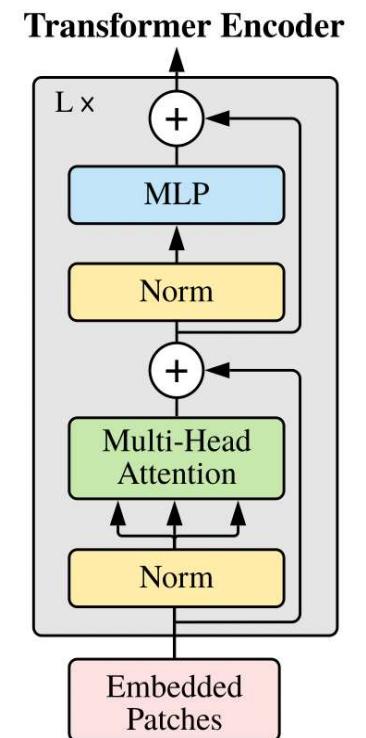
Ekstra Öğrenilebilir Sınıf Yerleştirme

ViT, NLP'deki [CLS] tokenine benzer şekilde, dizinin başına ekstra bir öğrenilebilir sınıf yerleştirme tokeni ekler. Bu token, tüm parçaların temsillerini birleştirerek görüntünün genel bir temsilini öğrenir ve sınıflandırma görevi için kullanılır. Bu, modelin tüm parçalar arasındaki ilişkileri dikkate alarak sınıflandırma yapmasını sağlar.

ViT Mimarisi

Transformatör Encoder

Transformatör encoder, parçaların vektör temsillerini (sınıf yerleştirme tokeni dahil) alır ve dikkat mekanizması kullanarak parçalar arasındaki ilişkileri modelleyerek anlamlı temsiller üretir. Encoder, birden fazla katmandan oluşabilir ve her katmanda parçalar, konumsal kodlama ve dikkat mekanizması uygulanır. Dikkat mekanizması, her bir parçanın diğer parçalara ne kadar "dikkat ettiğini" gösteren ağırlıklar üretir ve bu sayede model, parçalar arasındaki uzun mesafeli bağımlılıkları öğrenebilir. Encoder, her bir parçanın temsilini, diğer tüm parçaların temsilleriyle karşılaştırarak günceller. Bu, modelin görüntünün farklı bölgeleri arasındaki ilişkileri öğrenmesini sağlar.



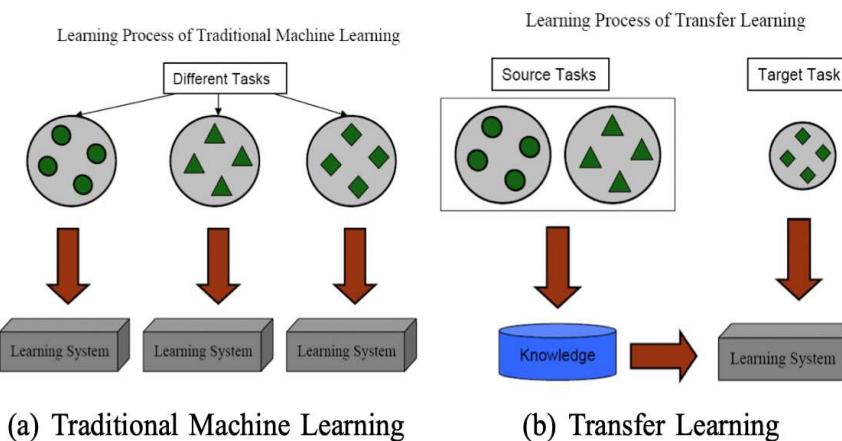
ViT Mimarisi

MLP Head (Çok Katmanlı Perceptron Başı)

Encoder tarafından üretilen temsiller, sınıflandırma veya diğer görevler için MLP head katmanına gönderilir. MLP head, genellikle tam bağlı katmanlardan oluşur ve nihai tahminleri yapar. MLP head, encoder tarafından üretilen temsilleri sınıflandırma skorlarına veya diğer çıktı formatlarına dönüştürür. Örneğin, görüntü sınıflandırma görevinde, MLP head, her bir sınıf için bir olasılık değeri üretir.

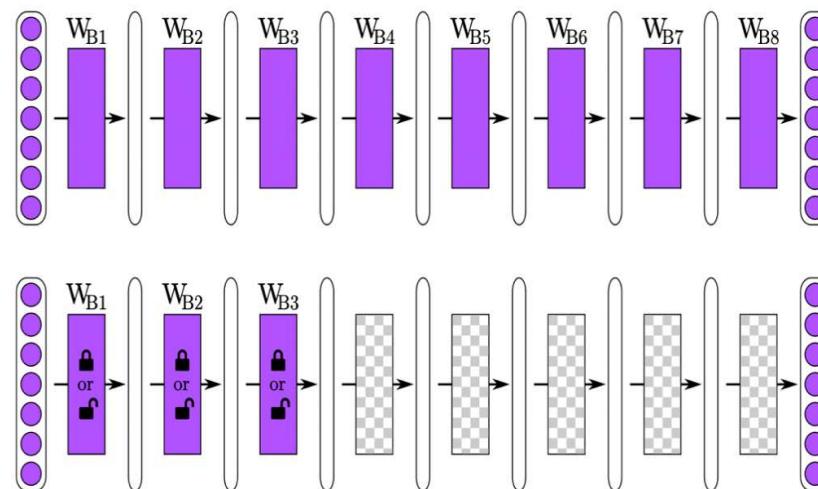
Transfer Öğrenme

Transfer öğrenme, önceden eğitilmiş bir modelin bilgisini ve ağırlıklarını yeni bir görevde aktarmak için kullanılan bir tekniktir. Bu teknik, özellikle sınırlı veri kümeleri için model eğitimini hızlandırır ve daha iyi sonuçlar elde edilmesini sağlar. Transfer öğrenme, genellikle derin öğrenme modellerinde kullanılır ve önceden eğitilmiş modeller, büyük veri kümeleri üzerinde eğitilerek genel özellikler öğrenir. Bu özellikler, daha sonra yeni görevde özgür verilerle ince ayar (fine-tuning) yapılarak modele aktarılır.



Fine-Tuning

Fine-tuning, önceden eğitilmiş bir modelin ağırlıklarını yeni görevde özgü verilerle ince ayar yapma işlemidir. Bu işlem, modelin yeni görevde uyum sağlamasını ve daha iyi sonuçlar elde edilmesini sağlar. Fine-tuning, genellikle modelin son katmanlarındaki ağırlıkları güncelleterek yapılır, ancak daha derin katmanlardaki ağırlıklar da ince ayarlanabilir. Fine-tuning, transfer öğrenmenin önemli bir parçasıdır ve modelin performansını önemli ölçüde artırabilir.



Derin Öğrenme ve Uydu Görüntüleri: Yangın Tespiti İçin Güçlü Bir Birleşim

Uydu Görüntüleri ve Derin Öğrenme

Günümüzde, orman yangınlarının tespiti ve izlenmesinde uydu görüntüleri ve derin öğrenme algoritmaları büyük bir potansiyel sunmaktadır. Uydu görüntüleri, geniş coğrafi alanların düzenli ve güncel bir şekilde izlenmesini sağlayarak geleneksel yöntemlere göre önemli avantajlar sunmaktadır. Derin öğrenme modelleri ise, bu görüntülerdeki karmaşık kalıpları analiz ederek yangınları otomatik olarak tespit etme yeteneğine sahiptir.

Derin Öğrenmenin Faydaları

Derin öğrenme, orman yangınlarının tespitinde birçok fayda sağlamaktadır: Daha hızlı tespit, daha geniş kapsama, daha düşük maliyet ve sürekli izleme imkanı sunar. Gerçek zamanlı veya neredeyse gerçek zamanlı yanık tespitine imkan tanır, geniş ve uzak coğrafi bölgeleri izleyebilir, geleneksel yöntemlere göre daha düşük maliyetli olabilir ve hava koşullarından bağımsız olarak 7/24 izleme sağlar.

Vision Transformer (ViT): Görüntü Analizinde Yeni Bir Çağ

ViT Mimarisi

Başlangıçta doğal dil işleme için geliştirilen Vision Transformer (ViT), dikkat mekanizmalarını kullanarak görüntülerdeki karmaşık kalıpları anlamada olağanüstü yetenekler sergilemiştir. Geleneksel CNN'lerin aksine, ViT'ler görüntüyü bir dizi yama olarak ele alır ve her bir yamayı birer kelime gibi işleyerek aralarındaki ilişkileri anlamaya çalışır.

Dikkat Mekanizmasının Gücü

ViT'nin en önemli özelliği, görüntülerdeki uzun menzilli bağımlılıkları yakalayabilen dikkat mekanizmasıdır. Bu, yangınların farklı şekil ve boyutlarda olabileceği ve görüntünün farklı bölgelerine dağılabileceği orman yangını tespiti gibi uygulamalar için idealdir. ViT, duman, alev, renk değişiklikleri ve yangınlarla ilişkili diğer görsel ipuçlarını etkili bir şekilde tespit ederek analiz edebilir.

Çalışmanın Temeli

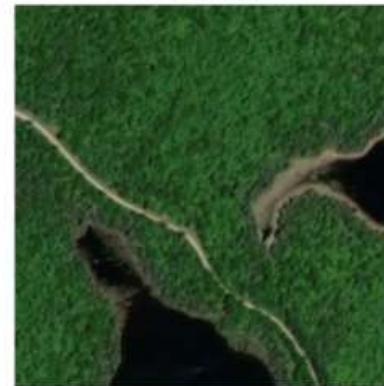
0: nowildfire



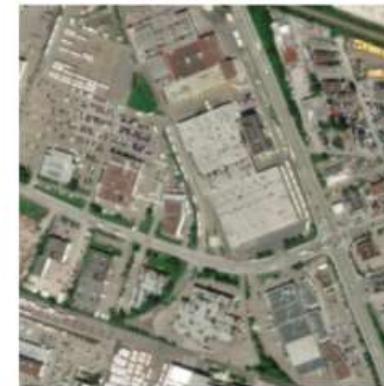
0: nowildfire



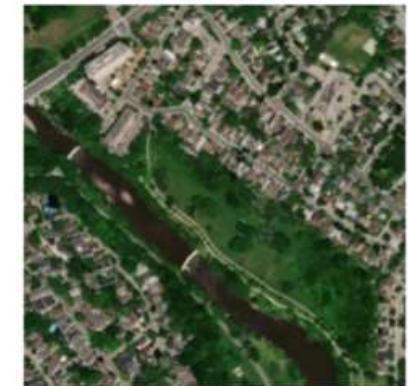
1: wildfire



0: nowildfire



0: nowildfire



Veri Seti

Kanada Orman Yangınları Veri Seti

Kaynak

Kanada Hükümeti Açık Veri Portalı

Lisans

Creative Commons 4.0 Atif (CC-BY)

Veri Seti Açıklaması

Veri seti, Kanada'nın farklı bölgelerinde çekilmiş 350x350 piksel boyutlarında uydu görüntülerinden oluşmaktadır.

Her görüntü, iki sınıftan birine aittir:

Orman Yangın Riski: Yangın riski taşıyan bölgeleri gösteren görüntüler.

Yangın Riski Yok: Yangın riski taşımayan bölgeleri gösteren görüntüler.

Veri seti, toplamda 42.850 görüntü içermektedir ve bu görüntülerin dağılımı aşağıdaki gibidir:

Orman Yangın Riski: 22.710 görüntü

Yangın Riski Yok: 20.140 görüntü

Veri Hazırlığı: Modelleri Eğitmek İçin Verileri Hazırlama

-
- 1**
 - Veri Kümelerinin Bölünmesi
 - Eğitim, Doğrulama, Test
 - 2**
 - Eğitim Kümesi
 - %70 (30.250 görüntü)
 - 3**
 - Doğrulama Kümesi
 - %15 (6.300 görüntü)
 - 4**
 - Test Kümesi
 - %15 (6.300 görüntü)

Derin öğrenme modelini eğitmek ve değerlendirmek için veri seti üç ayrı kümeye ayrılmıştır: Eğitim Kümesi (%70), Doğrulama Kümesi (%15) ve Test Kümesi (%15).

- 
- 1**
 - Yeniden Boyutlandırma
 - ViT: 224x224 piksel
 - 2**
 - Diğer Modeller
 - 350x350 piksel
 - 3**
 - Normalleştirme
 - 0-1 arasında ölçeklendirme

Önişleme adımları olarak, görüntüler ViT mimarisinin gereksinimleri doğrultusunda 224x224 piksele yeniden boyutlandırılmış, diğer modeller için ise orijinal 350x350 piksel çözünürlüğü korunmuştur. Tüm görüntülerdeki piksel değerleri 0 ile 1 arasında ölçeklendirilmiştir.

Model Geliştirme: Araçlar ve Teknikler

Tüm modeller, Metal Performance Shaders (MPS) arka ucu kullanılarak bir Macbook Air M2 bilgisayarda eğitildi. MPS, Apple silikonu için optimize edilmiş bir makine öğrenimi kitabıdır ve bu da model eğitimini önemli ölçüde hızlandırır. Model geliştirme ve eğitim için TensorFlow ve PyTorch derin öğrenme kitaplıkları kullanıldı.

Genel teknikler olarak, modelin eğitim veri setine aşırı uymasını önlemek için erken durdurma tekniği kullanıldı. Doğrulama kaybının belirli bir süre boyunca iyileşmediği noktada model eğitimini durdurarak, daha önce görülmemiş verilere daha iyi genelleme yapılmasını sağlar. Ayrıca, model eğitimi sırasında, belirli aralıklarla kontrol noktaları kaydedildi. Kontrol noktaları, modelin ağırlıklarının kaydedilmiş anlık görüntüleridir ve eğitim sırasında herhangi bir kesinti olması durumunda eğitime devam etmek için kullanılabilir.

Model Geliştirme: Karşılaştırılan Modeller

Özel Evrişimli Sinir Ağı (CNN)

Orman yangını tespiti için özel olarak tasarlanmış bir CNN modeli.

EfficientNetB4

Doğruluk ve verimlilik için optimize edilmiş ölçeklenebilir bir CNN mimarisi.

Vision Transformer (ViT)

Görüntüleri bir dizi yama olarak işleyen ve aralarındaki ilişkileri dikkat mekanizmalarıyla analiz eden yeni nesil bir derin öğrenme modeli.

ResNet50

Atlama bağlantıları kullanan önceden eğitilmiş ve yaygın olarak kullanılan 50 katmanlı bir derin CNN.

EfficientNetB7

EfficientNet ailesinin daha derin ve daha geniş bir versiyonu.

Önceden Eğitilmiş ViT

Transfer öğrenme için ImageNet veri seti üzerinde önceden eğitilmiş bir ViT modeli ("google/vit-base-patch16-224-in21k").

Model Eğitimi: Ortak Hiperparametreler

| Hiperparametre | Değer |
|--------------------------|---------------------|
| Epoch Sayısı | 10 |
| Batch Boyutu | 32 |
| Optimizasyon Algoritması | Adam |
| Kayıp Fonksiyonu | binary_crossentropy |
| Veri Normalizasyonu | Rescale 1./255 |

Model Detayları: Özel Tasarlanmış CNN

Özel olarak tasarlanmış bir Evrişimli Sinir Ağrı (CNN) modeli, orman yangını tespiti için kullanılmıştır. Bu model, yangınların tespitinde önemli özellikler çıkarmak üzere optimize edilmiştir.

Mimarisi

Modelin mimarisi aşağıdaki bileşenlerden oluşmaktadır:

- 3 Evrişim Katmanı:
Özellik haritaları çıkarmak için 32, 64 ve 128 filtre kullanır.
- 2 Maksimum Havuzlama Katmanı:
Boyutsallığı azaltmak ve en önemli özellikleri korumak için.
- 2 Tam Bağlı Katman:
Sınıflandırma için.

Hiperparametreler

Modelin eğitiminde kullanılan temel hiperparametreler şunlardır:

- Aktivasyon Fonksiyonu:
ReLU (Evrişim ve Tam Bağlı Katmanlar)
- Çıkış Katmanı Aktivasyonu:
Sigmoid
- Dropout Oranı:
0.5 (Aşırı uydurmayı önlemek için)

Model Detayları: Önceden Eğitilmiş ResNet50

Atlama Bağlantılarının Önemi

ResNet50, atlama bağlantıları kullanarak derinliğin getirdiği zorlukları aşar. Atlama bağlantıları, bilginin ağda daha serbestçe akmasını sağlayarak kaybolan gradyan problemini azaltır ve daha derin ağların eğitimini mümkün kılar.

Hiperparametreler

- Önceden Eğitilmiş Ağırlıklar: ImageNet
- Dondurulmuş Katmanlar:
İlk katmanlar donduruldu, böylece önceden öğrenilen özellikler korunarak modelin orman yangını tespiti görevi için daha hızlı ve etkili bir şekilde ince ayarı sağlandı.
- Çözülen Katmanlar:
Son 4 katman, orman yangını tespit görevine uyum sağlamak için ince ayar yapıldı.
- Diğer Hiperparametreler:
Ortak Hiperparametreler slaytındaki gibi.

Model Detayları: Verimli ve Doğru - EfficientNet

EfficientNet'in Ölçeklendirme Yöntemi

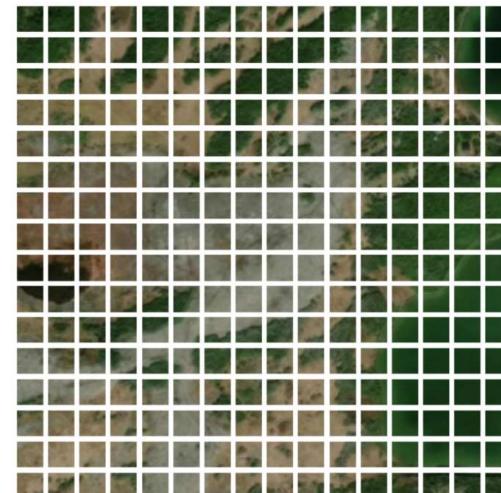
EfficientNet, bir CNN'in derinliğini, genişliğini ve çözünürlüğünü dengeli bir şekilde ölçeklendirerek daha iyi performans elde eden yeni bir model ölçeklendirme yöntemi kullanır. Bu dengeli ölçeklendirme yaklaşımı, daha az parametre ve hesaplama gücü kullanarak daha yüksek doğruluk elde edilmesini sağlar.

Hiperparametreler

- Önceden Eğitilmiş Ağırlıklar: ImageNet
- Dondurulmuş Katmanlar: İlk katmanlar
- Çözülen Katmanlar: Son 4 katman
- Diğer Hiperparametreler:
Ortak Hiperparametreler slaydındaki gibi.

Model Detayları: Vision Transformer - Yeni Nesil Görüntü Analizi

Vision Transformer (ViT), görüntülerini işlemek için yeni bir yaklaşım sunan yeni nesil bir derin öğrenme modelidir. ViT'in çalışma prensibi aşağıdaki adımlardan oluşur:



- 
- 1**
 - Yama Oluşturma
 - ViT, girdi görüntüsünü sabit boyutlu yamalara böler.
 - 2**
 - Doğrusal Projeksiyon
 - Her yama, düşük boyutlu bir vektöre düzleştirilir ve doğrusal bir projeksiyon katmanı kullanılarak yansıtılır.
 - 3**
 - Konum Kodlamaları
 - Her yamaya, modelin görüntü içindeki uzamsal konumunu korumasını sağlayan konum kodlamaları eklenir.
 - 4**
 - Transformatör Kodlayıcı
 - Düzleştirilmiş yama vektörleri, daha sonra kendi kendine dikkat mekanizmaları kullanarak farklı yamalar arasındaki ilişkileri öğrenen bir dizi Transformatör kodlayıcı katmanından geçirilir.
 - 5**
 - Sınıflandırma Token'ı
 - Sınıflandırıcı başlıktan geçen özel bir sınıflandırma token'inin çıktısı, nihai görüntü sınıflandırması için kullanılır.

ViT modelinin temel hiperparametreleri ise şu şekildedir:

- Yama Boyutu: 16x16 piksel
- Dikkat Başlığı Sayısı: 12
- Transformatör Katmanı Sayısı: 12

Model Detayları: Önceden Eğitilmiş ViT - Bilgi Birikimini Kullanma

Transfer Öğrenme ve İnce Ayar

Transfer öğrenme, daha önce büyük bir veri seti (ImageNet gibi) üzerinde eğitilmiş bir modelin, yeni bir görevde başlangıç noktası olarak kullanılmasıdır. İnce ayar ise, önceden eğitilmiş modeli yeni görevde daha iyi uyum sağlamak için yeni verilerle yeniden eğitmektir.

Kullanılan Model

"google/vit-base-patch16-224-in21k" - ImageNet-21k veri seti üzerinde önceden eğitilmiştir.

Hiperparametreler

- Dondurulmuş Katmanlar:
İlk 9 Transformatör katmanı donduruldu.
- Çözülen Katmanlar:
Son 3 Transformatör katmanı, katman normalleştirme katmanı ve sınıflandırıcı katmanı çözüldü ve orman yangını tespit görevi için ince ayar yapıldı.
- Dropout Oranı:
0.1 (Sınıflandırıcı katmanında)

Modellerin Değerlendirilmesi

CNN Modeli

ResNet50

EfficientNetB4

EfficientNetB7

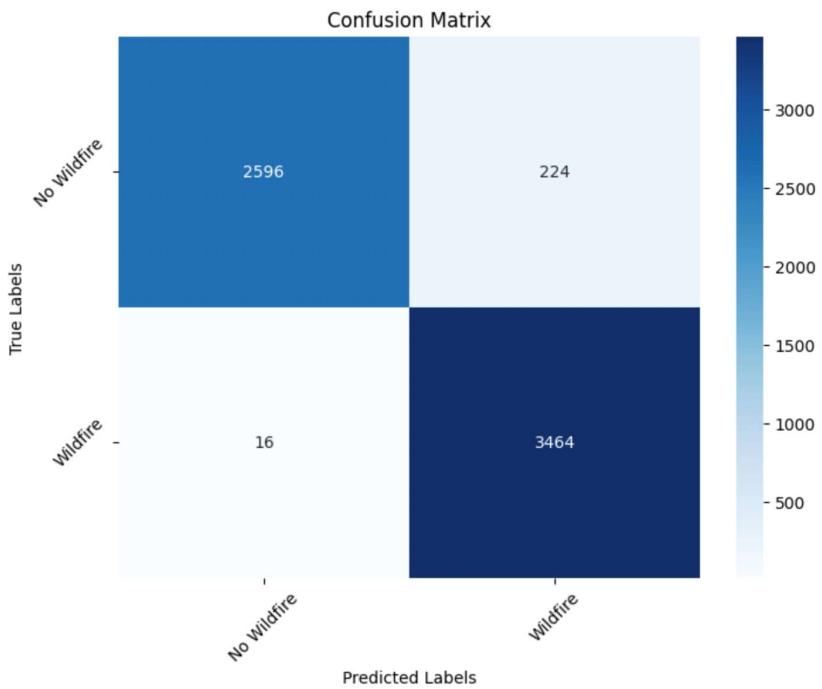
Vision Transformer
(ViT)

Önceden Eğitilmiş
Vision Transformer

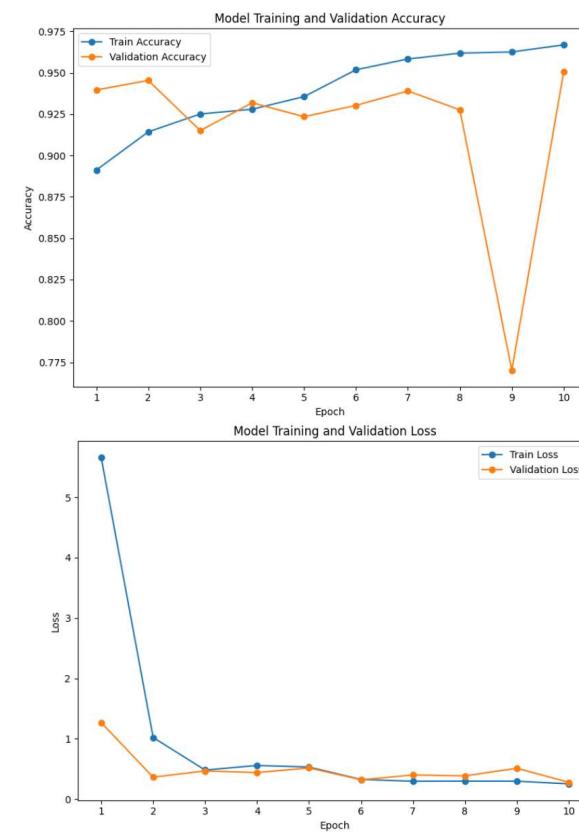
CNN Modeli

| CNN | Precision | Recall | F1 Skoru | Support |
|--------------|-----------|--------|----------|---------|
| No Wildfire | 0,99 | 0,92 | 0,96 | 2820 |
| Wildfire | 0,94 | 1,00 | 0,97 | 3480 |
| Accuracy | | | 0,96 | 6300 |
| Macro avg | 0,97 | 0,96 | 0,96 | 6300 |
| Weighted avg | 0,96 | 0,96 | 0,96 | 6300 |
| Doğruluk: | | 96,19% | | |

CNN Modeli



CNN Model Karışıklık Matrisi

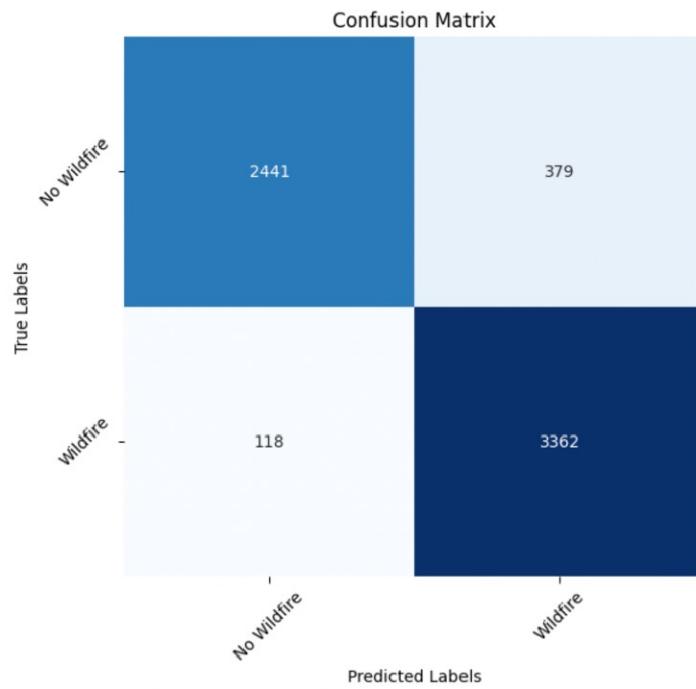


CNN Model Performans Raporu

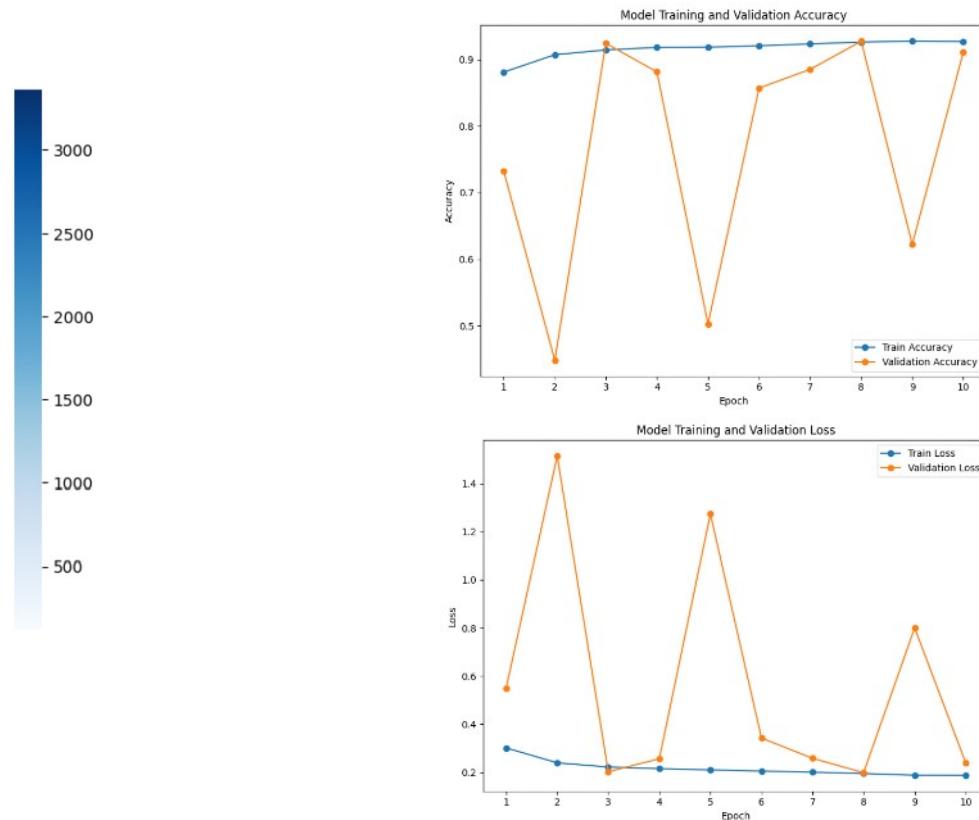
ResNet50

| ResNet50 | Precision | Recall | F1 Skoru | Support |
|--------------|-----------|--------|----------|---------|
| No Wildfire | 0,95 | 0,87 | 0,91 | 2820 |
| Wildfire | 0,90 | 0,97 | 0,93 | 3480 |
| Accuracy | | | 0,92 | 6300 |
| Macro avg | 0,93 | 0,92 | 0,92 | 6300 |
| Weighted avg | 0,92 | 0,92 | 0,92 | 6300 |
| Doğruluk: | | 92,11% | | |

ResNet50



ResNet50 Model Karışıklık Matrisi

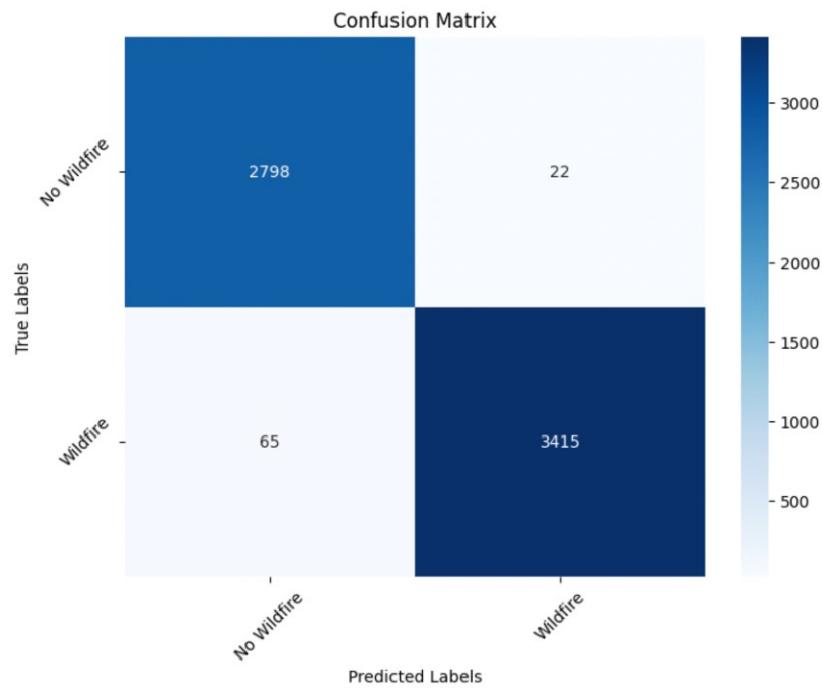


ResNet50 Model Performans Raporu

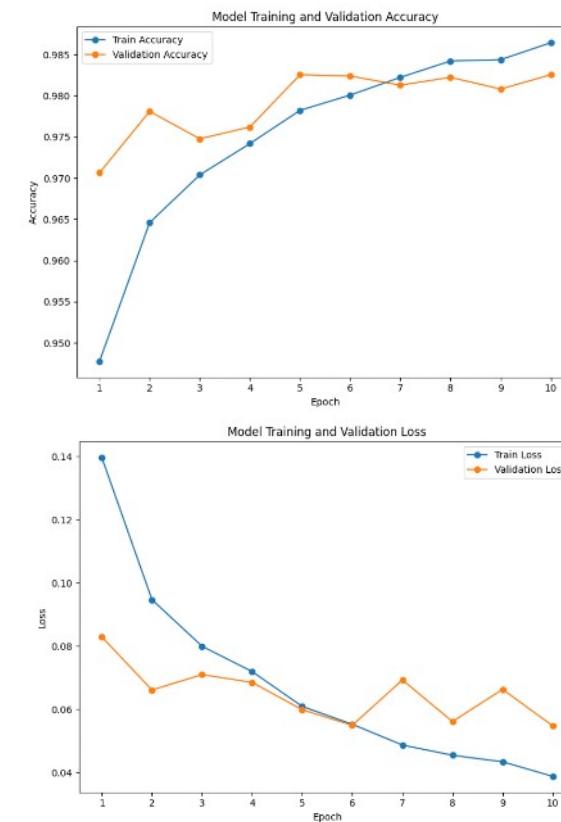
EfficientNetB4

| EfficientNetB4 | Precision | Recall | F1 Skoru | Support |
|----------------|-----------|--------|----------|---------|
| No Wildfire | 0,98 | 0,99 | 0,98 | 2820 |
| Wildfire | 0,99 | 0,98 | 0,99 | 3480 |
| Accuracy | | | 0,99 | 6300 |
| Macro avg | 0,99 | 0,99 | 0,99 | 6300 |
| Weighted avg | 0,99 | 0,99 | 0,99 | 6300 |
| Doğruluk: | | 98,62% | | |

EfficientNetB4



EfficientNetB4 Model Karışıklık Matrisi

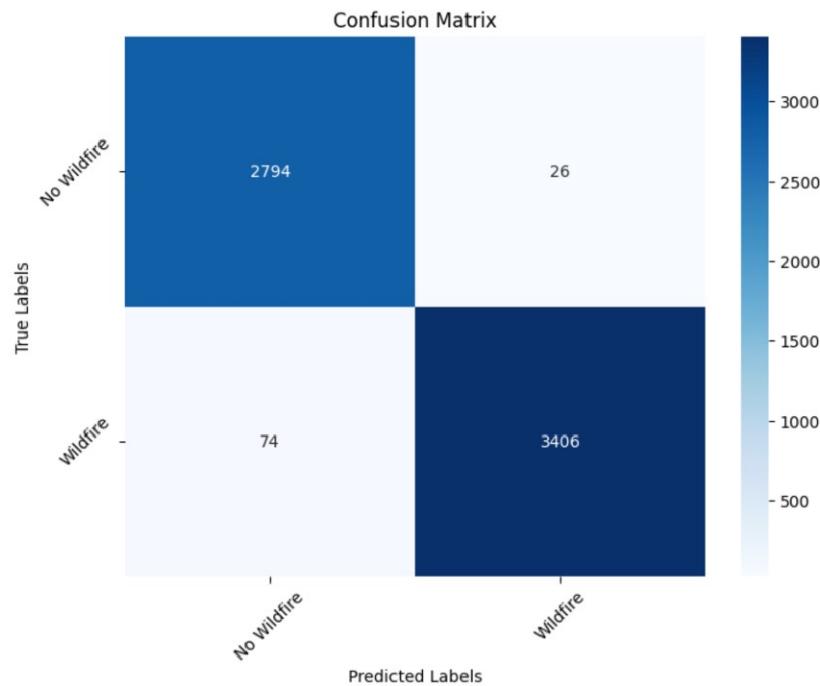


EfficientNetB4 Model Performans Raporu

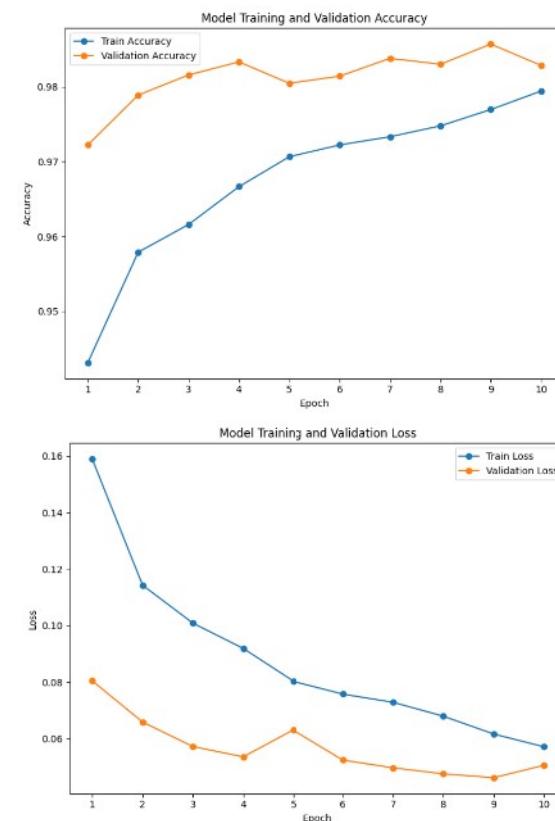
EfficientNetB7

| EfficientNetB7 | Precision | Recall | F1 Skoru | Support |
|----------------|-----------|--------|----------|---------|
| No Wildfire | 0,97 | 0,99 | 0,98 | 2820 |
| Wildfire | 0,99 | 0,98 | 0,99 | 3480 |
| Accuracy | | | 0,98 | 6300 |
| Macro avg | 0,98 | 0,98 | 0,98 | 6300 |
| Weighted avg | 0,98 | 0,98 | 0,98 | 6300 |
| Doğruluk: | | 98,41% | | |

EfficientNetB7



EfficientNetB7 Model Karışıklık Matrisi

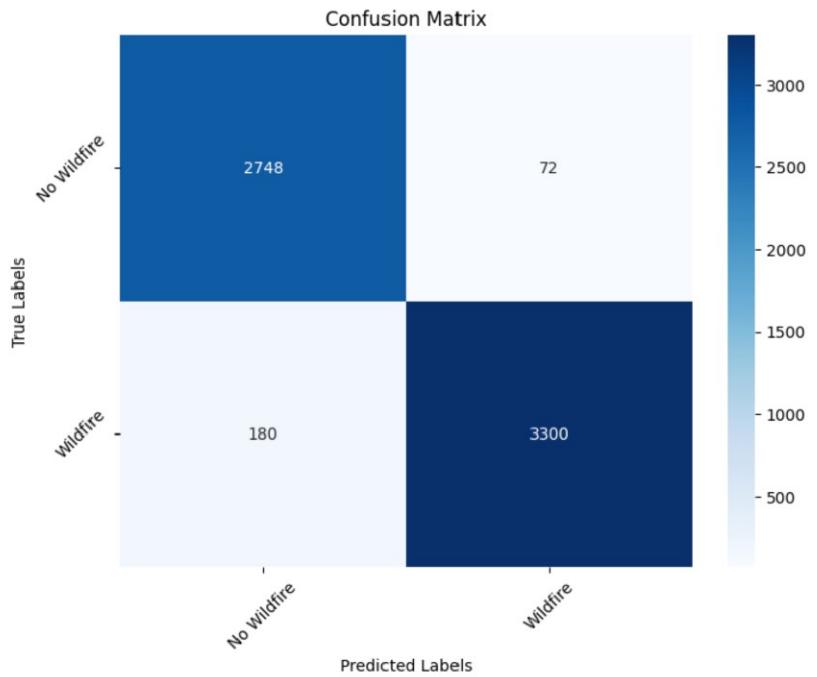


EfficientNetB7 Model Performans Raporu

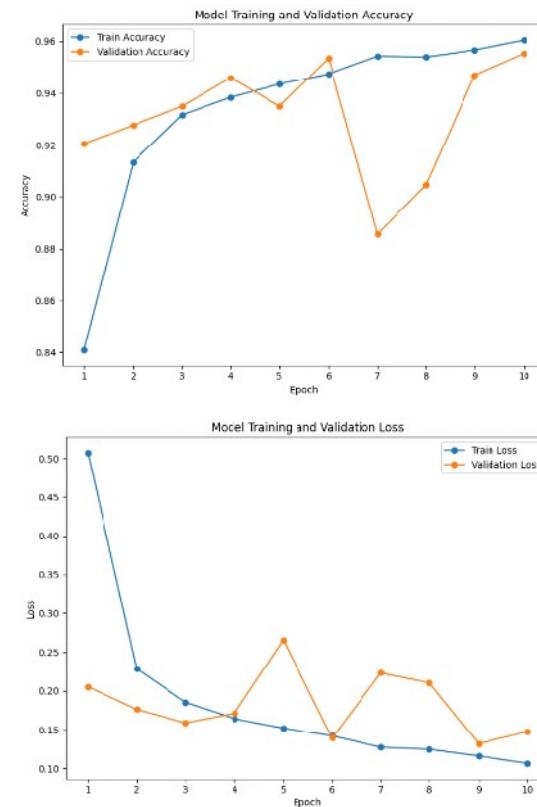
Vision Transformer (ViT)

| ViT | Precision | Recall | F1 Skoru | Support |
|--------------|-----------|--------|----------|---------|
| No Wildfire | 0,94 | 0,97 | 0,96 | 2820 |
| Wildfire | 0,98 | 0,95 | 0,96 | 3480 |
| Accuracy | | | 0,96 | 6300 |
| Macro avg | 0,96 | 0,96 | 0,96 | 6300 |
| Weighted avg | 0,96 | 0,96 | 0,96 | 6300 |
| Doğruluk: | | 96,00% | | |

Vision Transformer (ViT)



Vision Transformer (ViT) Model Karışıklık Matrisi

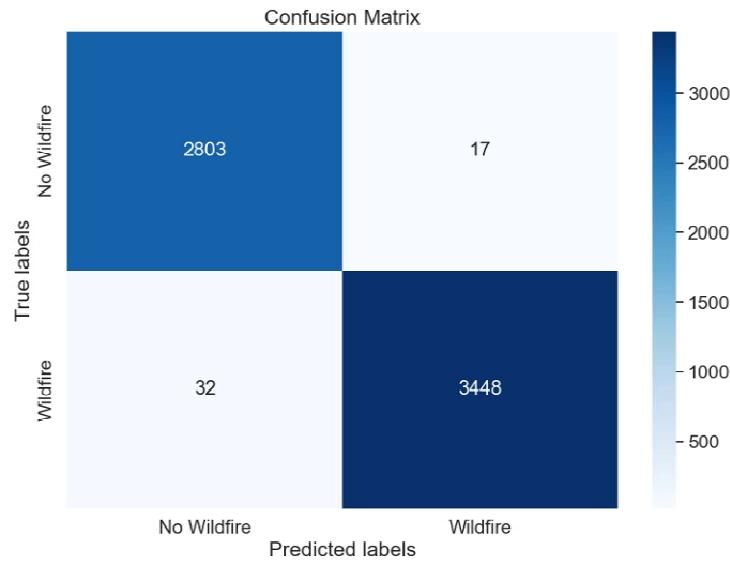


Vision Transformer (ViT) Model Performans Raporu

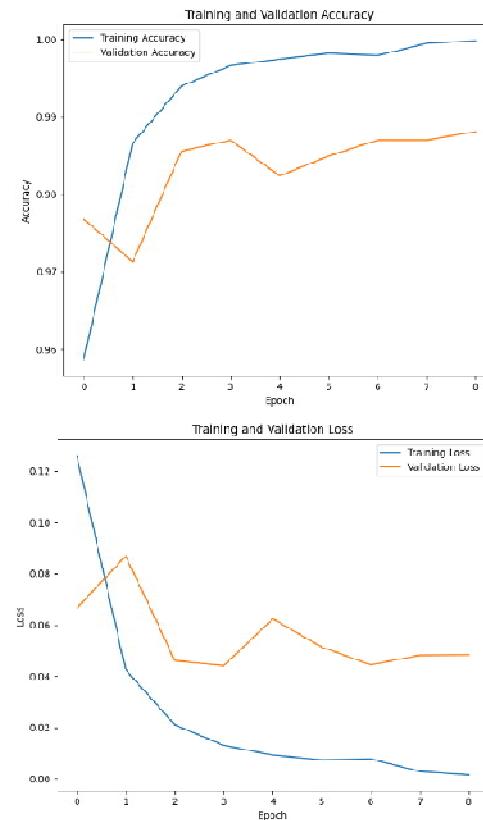
Önceden Eğitilmiş Vision Transformer

| Pretrained ViT | Precision | Recall | F1 Skoru | Support |
|----------------|-----------|--------|----------|---------|
| No Wildfire | 0,99 | 0,99 | 0,99 | 2820 |
| Wildfire | 1,00 | 0,99 | 0,99 | 3480 |
| Accuracy | | | 0,99 | 6300 |
| Macro avg | 0,99 | 0,99 | 0,99 | 6300 |
| Weighted avg | 0,99 | 0,99 | 0,99 | 6300 |
| Doğruluk: | | 99,22% | | |

Önceden Eğitilmiş Vision Transformer



Önceden Eğitilmiş Vision Transformer Model Karışıklık Matrisi



Önceden Eğitilmiş Vision Transformer Model Performans Raporu

Bulgular: Hangi Model Öne Çıktı?

Farklı derin öğrenme modellerinin orman yangını tespitindeki etkinliğini değerlendirmek için doğruluk, F1 skoru ve eğitim süresi gibi önemli ölçütler kullanıldı.

| Model | Eğitim Süresi (Saat) | Doğruluk (%) | F1 Skoru |
|-----------------------|----------------------|--------------|----------|
| Özel CNN | 2.47 | 96.19 | 0.97 |
| ResNet50 | 2.32 | 92.11 | 0.93 |
| EfficientNetB4 | 4.42 | 98.62 | 0.99 |
| EfficientNetB7 | 14.63 | 98.41 | 0.99 |
| ViT | 3.05 | 96.00 | 0.96 |
| Önceđen Eğitilmiş ViT | 4.42 | 99.22 | 0.99 |

Sonuçlar: ViT Neden Öne Çıkıyor?

Önceden eğitilmiş ViT modeli, %99.22'luk etkileyici bir doğruluk oranı ve 0.99'luk bir F1 skoru elde ederek diğer tüm modelleri geride bıraktı. ViT'nin dikkat mekanizması, görüntülerdeki uzun menzilli bağımlılıkları etkili bir şekilde yakalama yeteneğine sahiptir, bu da onu karmaşık sahnelerdeki yangınları tespit etmek için ideal hale getirir. Ayrıca, ViT'ler CNN'lere kıyasla daha az endüktif önyargıya sahiptir, bu da onları daha geniş bir yelpazedeği görüntü özelliklerine uyum sağlamanada daha esnek ve güçlü hale getirir. Son olarak, ViT'ler büyük veri kümeleri ve karmaşık görevlerle iyi ölçülebilir ve bu da onları gelecekteki orman yangını tespit sistemleri için umut verici bir teknoloji haline getirir.

Geleceğe Bakış: ViT ile Daha Güvenli Ormanlar

Ana Bulgular

- Bu çalışma, orman yangınlarının tespiti için Vision Transformer (ViT) modellerinin gücünü ve potansiyelini göstermektedir.
- Önceden eğitilmiş ViT modeli, test veri seti üzerinde %99,22'lük etkileyici bir doğruluk oranı elde ederek diğer derin öğrenme modellerinden daha üstün performans gösterdi.
- ViT'nin uzun menzilli bağımlılıkları yakalama, daha az endüktif önyargı ve ölçeklenebilirlik gibi avantajları, onu orman yangını tespiti için ideal bir aday haline getirmektedir.

Geleceğe Bakış: ViT ile Daha Güvenli Ormanlar

Gelecek Çalışmalar

- Farklı ViT mimarilerini (örneğin, Swin Transformer, DeiT) ve hiperparametre ayarlarını keşfetmek.
- Veri büyütme tekniklerini kullanarak modelin performansını ve genelleme yeteneğini artırmak.
- Modelleri gerçek zamanlı senaryolara ve farklı coğrafi konumlara uyarlamak.
- Farklı uydu sensörlerinden elde edilen verileri entegre ederek modelin doğruluğunu ve güvenilirliğini daha da artırmak.

Teşekkür

**Bu çalışma, Fırat Üniversitesi Bilimsel Araştırma Projeleri
Koordinasyon Birimi tarafından ADEP.23.08 proje numarası ile
desteklenmiştir.**

Dinlediğiniz İçin Teşekkür Ederim