



Metinden görsel oluşturma projesi

Mustafa Erdogdu

10.06.2024

Abstract

Bu projeyi oluşturmada temel amaç insanların hayal ettiği şeyleri muazzam uğraşlar vermeden görsel karşılıklarını alabilmelerini sağlamak. Bu hayal edilen şey belki bir rüya, belki bir senaryo veya geçmişte yaşadığı bir anı olabilir. Tabiki projenin ilk kısmı bu kadar detaylı görseller üretme konusunda yetersiz olsa da bu projenin belki de yapıtaşı olan “hayal etmek” kelimesi ilerleyen çalışmalarda böyle kompakt bir çalışma yapmak için fırsat tanıyor. Bu çalışmada yapılanları adım adım açıklayacak olursak başlangıç olarak görsel ayrıştırma modeli oluşturuldu daha sonra metin analiz modeli oluşturuldu. Bu iki model henüz birleştirilemedi ama metinden görsel oluşturma işlemi hazır görsel apileri kullanılarak sağlandı. Tasarlanan proje stable diffision modeline benzer şekilde üretilecekti ama projede ilerledikçe bu çizgiden biraz dışarı çıkıldı. stable diffisionda metinden görsel üretme işlemi gürültüden adım adım görüntü oluşturmak ile mümkün ama bu projede görsel oluşturmak için doğrudan metin vektörü kullanılması amaçlanıyor. Sonuçlar gösterdi ki

metinden görsel oluşturma projesi beklenilenden çok daha kompleks bir projeymiş. Kullanılan mimariler ve algoritmaları ortak bir projede birleştirmek için 10 haftalık proje süresi yeterli gelmedi. İlerleyen süreçte bu projeyi bitirmek için gerekli çalışmalara devam edilecektir.

1 Giriş

Bu projede, hazır kodlanmış Vgg16 modeli kullanılarak başka verileri ile bir model eğitimi gerçekleştirildi ve modele çıkarım katmanları eklenerek görsel ayrıştırma modeli oluşturuldu. GoogleNews-vectors-negative300.bin veri kümesi kullanılarak kelime benzerliği bulma yöntemiyle manzara kelimeleri keşfedildi ve Word2Vec(W2V) modeliyle benzerlikler analiz edildi. Bu proje için oluşturulan metin veritabanındaki promptlar kullanılarak bir W2V gömme modeli oluşturuldu. Ayrıca, webden metin kazıyarak kelime bulutları oluşturuldu ve kullanıcı tarafından girilen kelimeye benzer kelimelerin grafiği çıkarıldı. Bu kelime vektörleri görsel modellerle eşleştirilip metinden görsel oluşturma projesinde kullanılmak üzere kaydedildi. Kullanıcının girdiği prompt metin ön işleme adımlarıyla tek bir kelime vektörü haline getirildi ve Unsplash API'si kullanılarak bu promta göre görseller üretildi. Çalışmanın sonraki adımında Unsplash API yerine model.keras dosyası kullanılarak görseller üretilmesi hedefleniyor.

2 VERİ VE LİTERATÜR ARAŞTIRMASI

2.1 Nlp(doğal dil işleme)

Doğal dil işleme (NLP), bilgisayarlara insan dilini yorumlama, işleme ve anlama yeteneği veren bir makine öğrenimi teknolojisidir. insan dilini işlemek için hesaplamalı dil bilim, makine öğrenimi ve derin öğrenme modellerini birleştirir. NLP yazılımları, çeşitli uygulamalar için verileri hazırlamak adına belirteçlere ayırma, kök ayırma, kök çözümleme ve etkisiz kelimelerin kaldırılması gibi ön işleme tekniklerini kullanır.

2.2 Dikkate dayalı modelleme

Dikkate dayalı modeller , bir girdinin en önemli parçalarına odaklanmak ve bunları işlemek için kullanılan bir dizi model türüdür. Bu modeller, girdinin tümünü işlemek yerine, hangi parçaların en alakalı ve bilgilendirici olduğunu belirlemek için bir dikkat mekanizması kullanır.[1]

2.2.1 Transformatörler nedir

'Attention Is All You Need ' başlıklı makale transformatörleri tanıtıyor. Başlığın da belirttiği gibi, daha önce gördüğümüz dikkat mekanizmasını kullanıyor.[2] kelimeler arasındaki uzun mesafeli bağımlılıkları modellemek için kullanılan bir tür dikkat mekanizmasıdır. Transformatörler, bu mekanizmayı, bir metnin veya kodun farklı parçaları arasındaki ilişkileri öğrenmek için kullanır. Transformer iki parça (Encoder ve Decoder) yardımıyla bir diziyi diğerine dönüştürmeye yarayan bir mimaridir[3]

2.2.2 Gan

GAN, derin öğrenme alanında kullanılan bir yapay zeka (AI) modelidir. GAN, iki ayrı yapay sinir ağından oluşur

- Üretici (Generator): Bu ağ, girdi olarak rastgele gürültü alır ve gerçek verilere benzeyen yeni veriler (resimler, sesler, metinler vb.) üretmeye çalışır.
- Ayırt Edici (Discriminator): Bu ağ, gerçek verileri üretici tarafından oluşturulan verilerden ayırt etmeye çalışır.

GAN'lar, birbirleriyle "çekişerek" öğrenirler. Üretici, ayırt ediciyi aldatmaya çalışarak daha gerçekçi veriler üretmeye çalışır. Ayırt edici ise, verilerin gerçek olup olmadığını daha iyi ayırt etmeye çalışır. Bu rekabet sayesinde, her iki ağ da zamanla gelişir ve daha iyi performans gösterir.[4]

2.3 Stable Video Diffüzyon ve Gan farklılıkları

- Stable Video Diffusion, gürültü difüzyon ve ters difüzyon modelleriyle çalışır.
- GAN, birbirleriyle rekabet eden üretici ve ayırt edici ağlardan oluşur.
- Stable Video Diffusion , GAN'lara kıyasla daha az hesaplama gücü gerektirir ve bu nedenle daha hızlı çalışır. •Her iki model de yapay zeka görüntü işlemede kullanılsa da, farklı çalışma prensiplerine sahiptir.
- Stable Diffusion: Portre resimleri oluşturma, arka planları değiştirme, nesne ekleme veya kaldırma gibi görüntü düzenleme uygulamalarında kullanılır.
- GANs: Daha çok yaratıcı görüntü oluşturma, nesnelerin 3 boyutlu modellerinin oluşturulması gibi alanlarda tercih edilir. [5]

2.4 Gail

GAIL, yapay zeka alanında kullanılan bir derin öğrenme tekniğidir. Bu teknik, diğerlerinden biraz farklı bir şekilde çalışan bir tür GAN olarak düşünülebilir. Gail spesifik konularda görsel oluşturma imkanı tanır .GAIL,

belli bir tarzda veya stildeki verileri üretmekte daha başarılı olabilir. Dezavantajı ise metinden görüntü oluşturmaması.

2.5 Gürültüden görüntü oluşumu

Metinden ses oluşturmak için kullanıcının verdiği komutlar önce şekil ?? deki gibi gürültü olarak oluşur. Sonrasında komutlar doğrultusunda veri tabanındaki görsellerle kıyas yaparak adım adım şekil ?? ,şekil ?? ve şekil ??'e doğru ilerleyerek görsel oluşumu sağlar.

2.6 doğal dil işleme(Nlp)

Nlp, bilgisayarlara insan dilini yorumlama, işleme ve anlama yeteneği veren bir makine öğrenimi teknolojisidir. insan dilini işlemek için hesaplamalı dil bilim, makine öğrenimi ve derin öğrenme modellerini birleştirir. NLP yazılımları, çeşitli uygulamalar için verileri hazırlamak adına belirteçlere ayırma, kök ayırma, kök çözümleme ve etkisiz kelimelerin kaldırılması gibi ön işleme tekniklerini kullanır.

2.7 Torch

Torch, bilimsel hesaplamalar yapmak için birçok araç sağlar. Özellikle tensörler üzerinde işlemler gerçekleştirmek için kullanılır. Tensörler, çok boyutlu diziler olarak düşünülebilir ve matrislere benzerler, ancak daha yüksek boyutlara sahip olabilirler.

2.8 TorchVision

Torchvision, PyTorch kütüphanesi için geliştirilmiş bir kütüphanedir. Görüntü işleme ve bilgisayarlı görü uygulamaları için önceden eğitilmiş modeller, veri kümeleri ve yaygın dönüşümler sunar. Bu da, bu tür uygulamaları geliştirmeyi ve prototiplemeyi büyük ölçüde kolaylaştırır.

Torchvision'un sunduğu bazı temel özellikler şunlardır:

Popüler Görüntü İşleme Modelleri: AlexNet, ResNet, VGGNet gibi önceden eğitilmiş ve kullanıma hazır görüntü işleme modelleri içerir.

Geniş Veri Kümesi Seçeneği: ImageNet, CIFAR-10, MNIST gibi yaygın görüntü veri kümelerine erişim sağlar.

Görüntü Dönüşümleri: Resim boyutlandırma, kırma, döndürme, renk uzay dönüştürme gibi yaygın görüntü dönüşümlerini kolayca uygulamanızı sağlar.

Ek Yardımcı Araçlar: Görselleştirme araçları, model optimizasyon araçları ve hata ayıklama araçları gibi ek yardımcı araçlar da sunar.

2.9 Gensim Kütüphanesi Ne İşe Yarar?

Gensim, doğal dil işleme (NLP) alanında kullanılan ücretsiz ve açık kaynaklı bir Python kütüphanesidir. Metin verisini analiz etmek ve işlemek için çeşitli araçlar ve algoritmalar sunar. Gensim'in temel işlevleri şunlardır:

Konu Modelleme:

Büyük metin kümelerini analiz ederek gizli temaları ve konuları otomatik olarak belirler. Farklı metin türleri için uygun konu modelleme algoritmaları sunar.

Kelime Vektörleme:

Kelimeleri anlamsal ilişkilerine göre sayısal vektörlere dönüştürür. Bu vektörler, kelimelerin anlamlarını ve ilişkilerini temsil etmek için kullanılabilir.

Benzerlik Bulma:

Metinler, kelimeler ve cümleler arasındaki benzerliği hesaplar. Bu bilgiler, metin sınıflandırma, özetleme ve öneriler gibi görevlerde kullanılabilir.

Sözlük İşleme:

Metinlerden kelime ve kelime öbeklerini otomatik olarak çıkarır ve bunları işler. Bu bilgiler, duygu analizi, anahtar kelime çıkarma ve metin özetleme gibi görevlerde kullanılabilir.

Anlamsal Analiz:

Metinlerin anlamlarını ve ilişkilerini analiz etmek için çeşitli araçlar ve algoritmalar sunar. Bu bilgiler, metin sınıflandırma, duygu analizi ve metin özetleme gibi görevlerde kullanılabilir.

Gensim'in Kullanım Alanları: Haber ve blog yazılarını analiz etme Sosyal medya verilerini işleme Müşteri yorumlarını analiz etme Bilimsel makaleleri ve raporları analiz etme Otomatik metin özetleme ve çeviri Sohbet robotları ve sanal asistanlar geliştirme Bilgi arama ve öneri sistemleri oluşturma

Gensim'in Avantajları:

Kullanımı kolay ve öğrenmesi hızlıdır. Çok çeşitli NLP görevlerini destekler. Etkili ve ölçeklenebilir algoritmalar sunar. Geniş bir topluluk tarafından desteklenir.

Gensim'in Dezavantajları:

Bazı gelişmiş NLP görevleri için daha özel kütüphanelere ihtiyaç duyulabilir. Büyük metin kümelerini işlemek için zaman ve kaynak gerektirebilir.

Gensim'i Kullanmaya Başlamak: Gensim'i kullanmak için Python'u ve NumPy gibi bazı temel kütüphaneleri yüklemeniz gerekir. Gensim'i pip kullanılarak kurabilirsiniz: "pip install gensim" Daha fazla bilgi için Gensim'in resmi belgelerine bakabilirsiniz[6]. Özetle, Gensim, metin verisini analiz et-

mek ve işlemek için güçlü ve çok yönlü bir araçtır. NLP alanında çalışanlar veya metin verisinden bilgi çıkarmak isteyenler için idealdir.

2.10 Gensim’de Kelime Vektörlere Nasıl Yapılır?

Gensim kütüphanesi, kelimeleri anlamsal ilişkilerine göre sayısal vektörlere dönüştürmek için çeşitli kelime vektörlere algoritmaları sunar.

3 YÖNTEM

Çalışmada kullanılan yöntem ve modeller

3.1 Stable Video Diffusion nedir(SVD)?

SVD, yapay zeka (AI) alanında bulunan ve metinsel açıklamalara dayanarak video oluşturmak için kullanılan bir modeldir. 2022 yılında piyasaya sürülen bu model, kısa sürede ilgi çekmiş ve video oluşturma alanında yeni ufuklar açmıştır.[7]

3.1.1 Stable Video Diffuzyon ile Metinden Görsel Oluşturma Adımları

- Oluşturmak istediğiniz görseli detaylı ve net bir şekilde anlatan bir metinsel açıklama hazırlayın.
- Açıklamada renkler, nesneler, arka plan, kompozisyon ve görsel stil gibi unsurları detaylandırmaya çalışın.
- Örneğin, ”Bir deniz kenarında, gün batımında, palmiye ağaçlarının altında duran bir çift sallanan sandalye” gibi bir açıklama hazırlayabilirsiniz.
- Metinsel açıklamayı Stable Video Diffusion modelinin kabul ettiği bir formata dönüştürmek için bir metinden vektöre (text-to-vector) modeli kullanılır.CLIP veya VQ-VAE gibi metinden vektöre modeller, metinsel açıklamayı sayısal bir vektöre dönüştürerek modelin anlayabileceği bir formata dönüştürür.
- Model, metinsel vektörü kullanarak rastgele gürültüden başlayarak gerçekçi bir görüntü oluşturmaya başlayacaktır.[7]

3.2 Automatic1111 nedir

Automatic1111, Stable Diffusion isimli metinden görsele dönüştüren güçlü bir yapay zeka modelini kullanmak için kullanılan popüler bir grafik arayüzüdür (GUI). Stable Diffusion ile etkileşim kurmanızı kolaylaştırarak, metin açıklamalarına

veya prompt olarak bilinenlere dayalı olarak kolayca görüntü oluşturmaya olanak tanır.[8]

3.3 Civit ai nedir

Civit ai hazır eğitilmiş modelleri kullanmamıza olanak sağlayan sitedir. [9][10]

3.4 kullanılan kodlar

[11] numaralı kaggle çalışmasından faydalanıldı

3.5 Transformer Mimari

3.5.1 Transformer Mimari nedir

Transformer, Google AI tarafından 2017 yılında geliştirilmiş ve "Attention Is All You Need" isimli makalede sunulmuş bir derin öğrenme modelidir. Dikkat mekanizması ve kodlayıcı-çözücü mimarisi gibi özelliklerinden dolayı doğal dil işleme ve makine çevirisi gibi görevlerde oldukça başarılıdır. Transformer modeli, metinden görüntü oluşturmak, metinleri bir dilden diğerine çevirmek (Google Translate), metin üretmek (GPT-3), metin anlayışı (BERT) gibi birçok farklı görevde kullanılmaktadır. Dikkat mekanizması sayesinde uzun vadeli bağımlılıkları modelleyebilme, paralel işleme için uygun bir mimariye sahip olma ve farklı görevler için uyarlanabilme gibi avantajları vardır. Fakat RNN'lere kıyasla daha karmaşık bir mimariye sahip olması ve eğitimi için büyük veri kümeleri ve güçlü işlem gücü gerekli olması gibi dezavantajları da mevcuttur.[2]

3.5.2 Transformer Mimari katmanları

Şekil 1'de transformer mimari katmanları görülmektedir

Embedding Layer: Bu katman kelimeleri veya diğer girdi simgelerini gerçek sayı vektörlerine dönüştürür.

Positional Encoding: Transformatör mimarisi sıralı verileri işlemek için konumsal kodlama kullanır. Konumsal kodlama, dizideki her bir belirtecin konumunu belirten ek bilgidir.

Attention Layer: Dikkat katmanı, girdi dizisinin hangi bölümlerinin çıktıyı oluşturmak için en önemli olduğunu belirler.

Feed-Forward Network (FFN): FFN, dikkat katmanından gelen bilgileri işler ve çıktı için bir tahmin üretir.

Encoder and Decoder: Transformatör mimarisi çoklu kodlayıcı ve kod

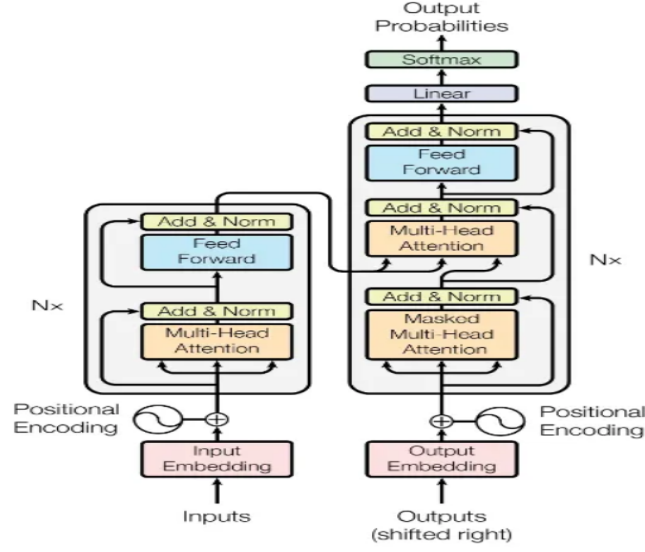


Figure 1: The Transformer - model architecture.

Şekil 1: trasnformer Mimari

çözücü katmanlarından oluşur. Her katman, giriş dizisini daha iyi işlemek ve daha doğru bir çıktı üretmek için bir önceki katmanın çıktısını kullanır.[11]

3.6 Transformer Mimari nasıl çalışır

“Input embedding” ile her satırı bir kelimeye karşılık gelen bir matrise çevirilecek . Sonra bu matris “encoder ” bloğuna girecek ve “multihead attention”, biriminden geçerek ve yeni bir matris elde edilecek. Bu sırada “decoder(çözücü)” bloğuna bir adım önce yaptığımız görseller gelecek. Bu görseller yine matrise çevirilecek ve benzer işlemlere tabi tutulacak. Ancak bu sefer ikinci “multi-head attention” biriminde girdi olarak “encoder” bloğunun çıktısı alınacak. “Decoder” bloğunun çıktısı son bir lineer fonksiyondan geçecek ve buradan alınan çıktılar softmax fonksiyonundan geçerek sonraki görünrtü için bize tahmin verecek. Transformerler bu işlemlerde dikkat mekanizmasını kullanıyor.Dikkat mekanizmasını basitçe anlatmak gerekirse,bir girdi dizisinin hangi bölümlerinin bir görev için en önemli olduğunu belirlemek için kullanılan bir tekniktir. Bu mekanizma, bir modelin girdi dizisinin tüm bölümlerine eşit dikkat etmesi yerine, belirli bölümlere odaklanmasını sağlar.

3.7 Transformer Mimari avantajları ve dezavantajları

Transformer Mimarisi Avantajları:

- RNN'lere kıyasla daha uzun metin dizilerini daha iyi işleyebilir.
- Kelimeler arasındaki uzun mesafeli bağımlılıkları öğrenebilir.
- Daha hızlı ve daha verimlidir.
- Farklı NLP görevlerine uyarlanabilir.

Transformer Mimarisi Dezavantajları:

- RNN'lere kıyasla daha fazla veri ve işlem gücü gerektirir.
- Karmaşık bir mimaridir ve anlaması zor olabilir.

3.8 VGG16 Modeli Nedir?

VGG16, 2014 yılında Visual Geometry Group (VGG) tarafından geliştirilmiş bir konvolüsyonel sinir ağı (CNN) modelidir.

Özellikleri: Görüntü sınıflandırma, nesne algılama ve bölgeleme gibi birçok görevde oldukça başarılıdır.

16 katmandan oluşur ve bu katmanların 13'ü konvolüsyonel katmandır.

Her konvolüsyonel katmandan sonra bir max pooling katmanı bulunur.

Son katmanlarda ise tam bağlantılı katmanlar kullanılır. ImageNet veri kümesi üzerinde eğitilmiştir.

1000 farklı nesne kategorisini sınıflandırabilir.

3.9 Çalışmada geliştirilen modeller

3.10 Görüntü ayrıştırma modeli

Bu model veri setine ait manzara görsellerini ayrıştırma için kullanılmak üzere geliştirildi.

3.11 Word2Vec modeli

Metin analizi için geliştirilen Word2Vec, kelimeleri gerçek sayılardan oluşan vektörlere dönüştüren bir sinir ağı modelidir. Bu vektörler, kelimelerin anlamsal ilişkilerini ve benzerliklerini yakalamada oldukça etkilidir.

3.12 metin modeli oluřturma adımları

3.13 Metin kazıma(web scraping) nedir?

Metin kazıma , web sitelerinden veri toplama iřlemidir. Bu, bir web sayfasının HTML yapısını analiz ederek belirli verileri çıkarmayı ierir. Python’da metin kazıma iin en popler ktphanelerden bazıları BeautifulSoup, requests ve Selenium’dur. Ařaėıda, BeautifulSoup ve requests kullanarak basit bir metin kazıma rneėi verilmiřtir.Bu alıřmada BeautifulSoup kullanıldı.Bunun sebebi BeautifulSoup kk ve orta lekli projeler iin hızlı bir bařlangı sunar. Basit iřlemler iin ideal bir seimdir.

3.14 Kelime bulutları oluřturma?

WordCloud ile bir kelime bulutu oluřturmak, bir metin verisindeki kelimelerin sıklıėını grsel olarak temsil eden bir grafik yaratmak anlamına gelir. Bu grafik, daha sık geen kelimeleri daha byk ve daha belirgin bir řekilde gsterirken, daha az geen kelimeleri daha kk ve daha az belirgin bir řekilde gsterir. Bu, metindeki nemli veya baskın temaları hızlıca tanımlamak iin olduka yararlı bir tekniktir.

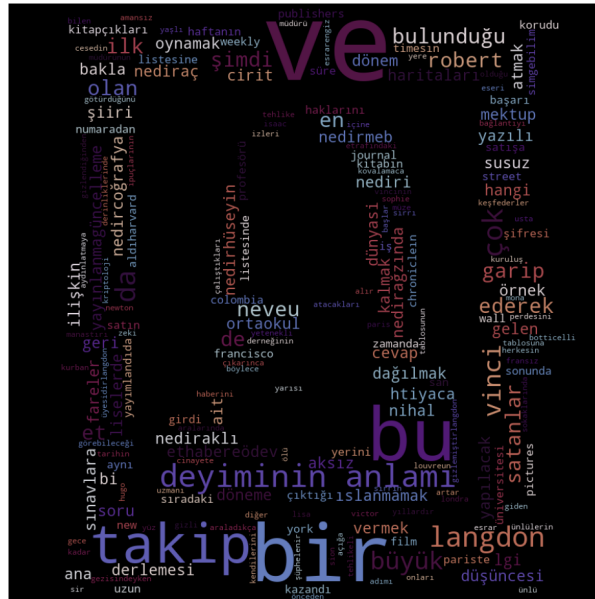
Kelime bulutu oluřturmak iin genellikle řu adımlar izlenir:

Metin Verisini Toplamak: Bir belge veya bir dizi metin verisi kullanılarak.

Metni Temizlemek ve n İřlemden Geirmek: Noktalama iřaretlerini kaldırmak, kk harfe dnřtrmek, stop words (nemsiz kelimeler) kaldırmak vb.

Kelime Sıklıėını Hesaplamak: Her kelimenin metinde ka kez getiėini saymak.

Kelime Bulutunu Oluřturmak: Bu sıklık bilgilerini kullanarak bir kelime bulutu grselleřtirmesi oluřturmak[?] .



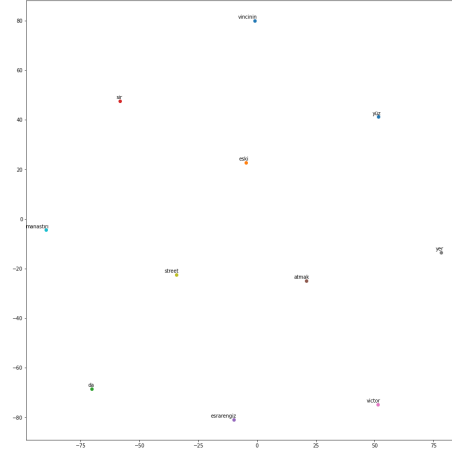
Şekil 2: kelime bulutu

3.15 Vektörlere ayırma

Metni bir vektörler koleksiyonuna dönüştürme işlemidir. Bu işlem, her kelimenin veya kelime öbeklerinin (örneğin, cümlelerin veya belgelerin) bir vektörle temsil edilmesini içerir. Metinleri vektörlere ayırmanın birkaç farklı yolu vardır, ancak en yaygın yaklaşımlardan biri, kelime gömme (word embedding) yöntemlerini kullanmaktır. Kelime gömme, kelimeleri belirli bir boyutta vektörlerle temsil etme işlemidir. Bu vektörler, kelimenin anlamını ve kullanım bağlamını yakalamak için eğitilmiş bir model tarafından öğrenilir. Örneğin, "köpek" kelimesi için bir kelime gömme modeli tarafından öğrenilen vektör şöyle olabilir: $\text{kopek}=[0.2,0.4,0.1,\dots,0.3]$ Bu vektörler, kelimenin anlamını bir sayısal formda temsil eder. Bu temsil, makine öğrenimi modellerinde ve doğal dil işleme uygulamalarında kullanılabilir. Örneğin, kelime vektörleri kullanılarak metin sınıflandırması, duygu analizi, kelime benzerlikleri hesaplama gibi birçok görev gerçekleştirilebilir. Eksi değerler, kelime vektöründeki belirli bir özelliğin, genel eğilimden veya diğer özelliklerden farklı olduğunu gösterebilir. Örneğin, bir kelimenin duygusal yükünü temsil eden bir vektörde, negatif bir bileşen, o kelimenin genel olarak olumsuz bir duyguyu ifade ettiğini gösterebilir.

3.16 Benzer kelime tespiti ve grafik oluřturma

Kelimeler arası benzerlik tespit edilirken kelime ok boyutlu bir vektör olarak temsil edilir. Bu vektörler, kelimelerin anlamlarını yakalar ve benzer anlamlara sahip kelimeler birbirine yakın vektörler olarak konumlandırılır. İki vektör arasındaki yakınlık veya benzerlik genellikle kosinüs benzerlięi ile ölçölür.



řekil 3: modelde satmak kelimesine en benzer 10 kelime

3.17 Word2Vec modellemesi nasıl yapılır?

Word2Vec, kelimeleri vektörlere dönüştürmek için kullanılan bir doğal dil işleme (NLP) modelidir. Kelimelerin anlamını sayısal bir forma çevirmek için iki temel yaklaşımdan birini kullanır: Continuous Bag of Words (CBOW) veya Skip-gram. Toplanan veriler öncelikle, veri ön işleme adımlarından geçer. Ardından, işlenmiş verilerle Word2Vec modeli eğitilir. Eğitilen model ile kelime vektörlerini inceleyebilir, kelimeler arasındaki benzerlikleri ve kelimeler arasındaki anlamsal ilişkileri analiz edebilirsiniz. Word2Vec modeli, kelimeleri anlamsal olarak anlamlandırmanın güçlü bir yoludur ve birçok NLP uygulamasında temel taşlardan biridir. Bu model Skip-gram algoritmasına göre kuruldu.

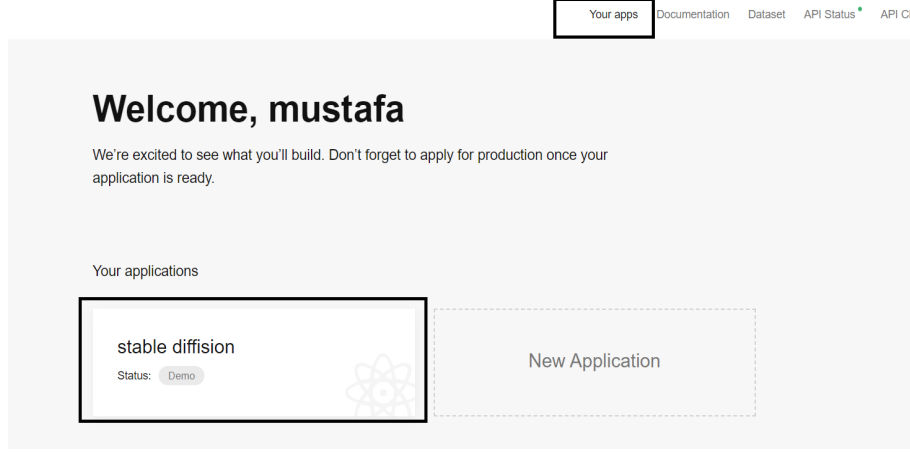
3.18 T-SNE (T-distributed Stochastic Neighbor Embedding) nedir?

T-SNE, yüksek boyutlu veri setlerini düşük boyutlu (genellikle 2 veya 3 boyutlu) bir alana indirgemek için kullanılan bir makine öğrenmesi algoritmasıdır. Bu yöntem, veriler arasındaki yerel benzerlikleri koruyarak veri görselleştirmesi için idealdir. Özellikle yüksek boyutlu kelime vektörlerini görselleştirmek için sıklıkla kullanılır. T-SNE'yi kullanarak Word2Vec modelindeki belirli bir kelimenin en yakın kelimelerini bulup bunların vektörlerini bulma gibi işlemlerde kullanılabilir.

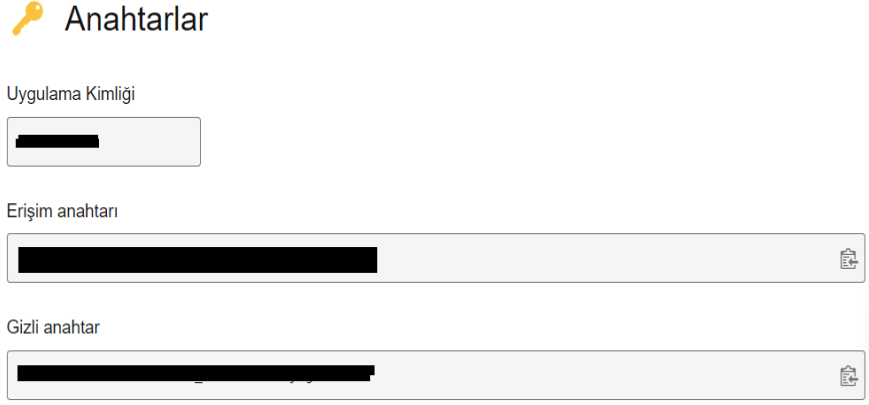
3.19 Unsplash API nedir?

Unsplash API, kullanıcıların Unsplash platformundaki yüksek kaliteli, telifsiz görsellere erişmesini ve bu görselleri uygulamalarında veya web sitelerinde kullanmalarını sağlayan bir programlama arayüzüdür. Unsplash, kullanıcılar tarafından yüklenen ve paylaşılan geniş bir fotoğraf koleksiyonuna sahiptir ve bu API, bu koleksiyona programlı erişim sağlar [12].

3.20 Unsplash API kullanımı ve proje oluşturma



Şekil 4: Unsplash app sayfası



Şekil 5: proje oluşturma

Yukarıdaki 5 numaralı şekilde görülen anahtarlar kodda gerekli kısımlara eklenerek projede kullanılabilir hale geliyor.

4 BULGU ve TARTIŞMA(hatalar burada)

Model eğitim değerleri

```
20/20 ————— 38s 2s/step - acc: 0.2308 - loss: 0.5242 - val_acc: 0.5000 - val_loss:
0.3920
Epoch 2/50
20/20 ————— 35s 2s/step - acc: 0.4379 - loss: 0.3875 - val_acc: 0.4500 - val_loss:
0.3524
```

Şekil 6: epoch değeri=50 epochun 1. ve 2. değerleri

```
Epoch 25/50
20/20 ————— 39s 2s/step - acc: 0.7319 - loss: 0.2009 - val_acc: 0.5500 - val_loss:
0.2433
Epoch 26/50
20/20 ————— 34s 2s/step - acc: 0.7179 - loss: 0.2069 - val_acc: 0.8500 - val_loss:
0.1790
```

Şekil 7: epoch değeri=50 epochun 25. ve 26. değerleri

```
Epoch 49/50
20/20 ————— 38s 2s/step - acc: 0.7763 - loss: 0.1728 - val_acc: 0.9000 - val_loss:
0.1187
Epoch 50/50
20/20 ————— 36s 2s/step - acc: 0.7519 - loss: 0.1956 - val_acc: 0.6000 - val_loss:
0.2349
```

Şekil 8: epoch değeri=50 epochun 49. ve 50. değerleri

Epoch: Bir epoch, modelin tüm eğitim setini bir kez gördüğü bir eğitim döngüsünü temsil eder. Görselde 50 epoch olduğunu görebilirsiniz, bu da modelin eğitim setini 50 kez gördüğü anlamına gelir.

Loss Loss, modelin tahminleri ile gerçek değerler arasındaki farkın bir ölçüsüdür. Düşük bir loss değeri, modelin daha iyi tahminler yaptığı anlamına gelir. Görselde loss değeri epoch boyunca azalmaktadır, bu da modelin eğitim ilerledikçe daha iyi hale geldiğini gösterir.

Accuracy: Accuracy, modelin doğru tahminlerde bulunma oranını temsil eder. Yüksek bir accuracy değeri, modelin daha iyi performans gösterdiği anlamına gelir. Görselde accuracy değeri epoch boyunca artmaktadır, bu da modelin eğitim ilerledikçe daha fazla doğru tahmin yaptığı gösterir.

Val loss: Val loss, doğrulama seti olarak adlandırılan ayrı bir veri kümesi üzerinde hesaplanan loss değeridir. Doğrulama seti, modelin eğitim setine aşırı uyum sağlamadığından emin olmak için kullanılır. Düşük bir val loss değeri, modelin genelleme yeteneğine sahip olduğunu gösterir, yani eğitim setinde görmediği veriler üzerinde de iyi performans gösterebileceği anlamına gelir. Görselde val loss değeri epoch boyunca azalmaktadır, bu da modelin genelleme yeteneğini geliştirdiğini gösterir.

Val accuracy: Val accuracy, doğrulama seti üzerinde hesaplanan accuracy değeridir. Yüksek bir val accuracy değeri, modelin genelleme yeteneğine sahip olduğunu gösterir. Görselde valaccuracy değeri epoch boyunca artmaktadır, bu da modelin genelleme yeteneğini geliştirdiğini gösterir.

4.0.1 Word2Vec modeli

Word2Vec'in iki ana modeli vardır:

CBOW (Continuous Bag-of-Words): Bu model, bir kelimenin etrafındaki kelimeleri tahmin etmek için kelimenin kendisini kullanır. Daha az veriyle daha hızlı bir şekilde eğitilmesi gereken modeller için daha uygundur.

Skip-gram: Bu model, bir kelimeyi tahmin etmek için etrafındaki kelimeleri kullanır. Word2Vec, kelime benzerliği bulma, ilgili kelimeler önerme ve metin sınıflandırma gibi çeşitli NLP görevlerinde kullanılabilir. Daha az gürültülü veriyle daha iyi sonuçlar veren modeller için daha uygundur.

4.0.2 GloVe(Global Vectors for Word Representation):

GloVe, kelimeleri hem yerel hem de küresel istatistikleri kullanarak vektörlere dönüştürür. Yerel istatistikler, kelimelerin kelime bağlam penceresinde birlikteliklerini temsil ederken, küresel istatistikler kelimelerin tüm metin kümesindeki birlikteliklerini temsil eder. GloVe, Word2Vec'e kıyasla daha az gürültülü ve

daha anlamlı vektörler üretme eğilimindedir. GloVe, kelime benzerliği bulma, duygu analizi ve makine çevirisi gibi çeşitli NLP görevlerinde kullanılabilir.

4.1 Veri kümesi ve kullanılan veriler

Veri kümesinin içeriği:Doğa görselleri

Veri kümesinin adı:Landscape - Pictures - litter on forest floor-dataset

Veri kaynağı:[13],[14],[15]

Veri örnekleri:



Şekil 9: Landscape-Pictures[13]

4.2 Kagglenden colaba veri çekme

Kaggle API'sini kullanarak bilgisayara jayson dosyası içerisinde token indiriyoruz.Burada muhtemel hata alma sebebi olarak bu jason dosyası birden fazla kez indirilirse uyumsuzluk oluyor ve hata alınır.Eğer jason dosyanızda hata alınmamışsa Colab'da veri çekme işlemi için öncelikle Şekil 10'deki Kaggle API anahtarını alıp ve Colab'da koddaki gerekli ayarlamaları yapmak gerekiyor.

erdogdu mustafa

Profile Settings

Full name
erdogdu

Avatar (optional)
Upload file

Homepage (optional)
Homepage

AI & ML interests (optional)
AI & ML interests

Github username (optional)
2118121024-mustafa-erdogdu

Twitter username (optional)
Twitter account

Şekil 10: kaggleToken

erdogdu mustafa

Profile Settings

Full name
erdogdu

Avatar (optional)
Upload file

Homepage (optional)
Homepage

AI & ML interests (optional)
AI & ML interests

Github username (optional)
2118121024-mustafa-erdogdu

Twitter username (optional)
Twitter account

Şekil 11: huggingFaceToken.png

4.3 Colab hugging face bağlama

Şekil 11'deki hugging face token anahtarını alıp ve Colab'da koddaki gerekli ayarlamaları yapmak gerekiyor.

```
!kaggle datasets download -d savasy/multiclass-classification-data-for-turkish-tc32
401 - Unauthorized - Unauthenticated
```

Şekil 12: webHata[?]

Colaba veri çekerken alınabilecek bazı hatalar ve çözümleri,Şekil 12'daki hata giriş yapılmak istenilen web sitesi adresinde giriş yetkilendirmesinin olmadığı belirtilen bir hatadır.hugging face ile colab e mailleri farklı olursa bu hata alınır. önüne geçmek için mailleri kontrol etmek gerekiyor.

4.4 Projede karşılaşılabilecek hatalar ve çözümleri

```
ValueError: Found input variables with inconsistent numbers of samples: [751, 4]
```

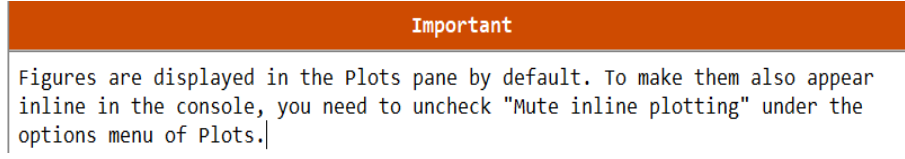
Şekil 13: hata1

Şekil 18 deki hatanın sebebi veri setindeki metinlerin hepsinin etiketli olmadığından veya etiket değerleri etiketlenmesi gereken veriden fazla olduğu zaman buhata alınır. Hata 14 de görecğimiz üzere word2vec modeli sınıflandırma

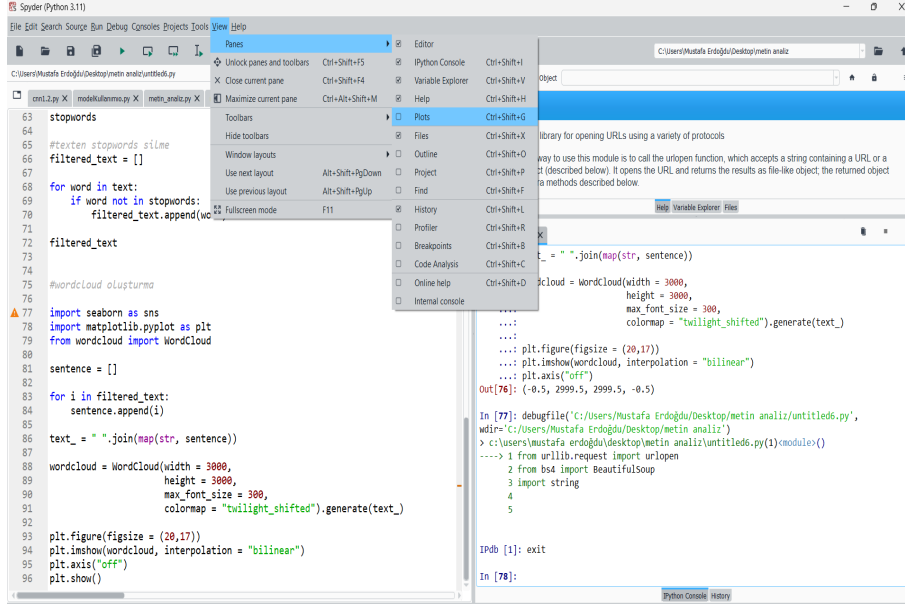
```
ValueError: This solver needs samples of at least 2 classes in the data, but the data contains only one class: 1
```

Şekil 14: hata2

modeli olduğu için en az iki sınıf olması gerekiyor.



Şekil 15: Görsel oluşturma uyarısı



Şekil 16: Grafik görüntüleme ayarı açılması

Şekil 15’de alınan uyarı 16 numaralı şekilde belirtildiği şekilde grafik görüntüleme ayarları açılır.

TypeError: Word2Vec._init__() got an unexpected keyword argument 'size'

Şekil 17: Size sürüm hatası.

şekil 17 ’de alınan hata size ifadesi w2w modelin eski sürümlerinde kullanılıyor.Yeni sürümde vektor size ifadesi kullanılması gerekiyor. Yukarıdaki 18 numaralı şekilde görülen hata erişim anahtarının hatalı girilmesinden kaynaklanıyor.Hatanın düzeltilmesi için "Unsplash API kullanımı ve proje oluşturma" adımını inceleyebilirsiniz.

```
Hata oluřtu: 401 Client Error: Unauthorized for url: https://api.unsplash.com/
search/photos?query=mountain+rang+snow-
cap+peak+offer+magnific+view&client_id=YOUR_UNSPASH_API_KEY
<Figure size 1080x720 with 0 Axes>
```

řekil 18: api unsplash hatası

5 SONUÇ

10 haftalık bir süreyi kapsayan bu projede, Stable Video Diffusion modelinin temel prensipleri ve kullandığı teknolojiler (örneğin, derin öğrenme, difüzyon modelleri) araştırılarak benzer teknolojiler ve çalışmalar (örneğin, DALL-E 2,gail) ile karşılaştırıldı. Bu sayede, Stable Video Diffusion modelinin güçlü ve zayıf yönleri belirlenerek, bu modele benzer bir modelin geliştirilmesine zemin hazırlandı.Stable diffision modelini farklı bir versiyonu geliştirilmeye çalışıldı. Proje kapsamında, modelin metin açıklamalarını doğru şekilde yorumlayabilmesi için gerekli veri setleri ve eğitim yöntemleri belirlendi, farklı manzara türlerini ve farklı metin formatlarını desteklemesi için gerekli geliřtirmeler yapıldı.

[16]

5.1 Gelecekte yapılması planlanan çalışmalar nelerdir?

Gelecek süreçte hedeflenen başarımlar, kullanıcının girdiği metinden metin vektörü elde edilip doğrudan bir manzara resmi oluşturabilen bir model inşa edilmiş olacak ve bu modelin farklı metin açıklamalarına ve parametrelere karşı test edilerek, modelin oluşturduğu resimlerin kalitesinin ve çeşitliliğinin değerlendirilmesi ve modelin sınırlamalarının belirlenmesi gibi işlemler tamamlanacaktır. Metinden görsel üretme işlemi stable diffision aksine doğrudan kelime vektörleri ile sağlanacak.

Kaynaça

- [1] J. Devlin, “Dil anlamak için derin Çift yönlü transformatörlerin Ön eğitimi erişim adresi.” <https://arxiv.org/abs/1810.04805>, (11 ekim 2023). Accessed: 10.03.2024.

- [2] A. Vaswani, “Attention is all you need.” [<https://arxiv.org/abs/1706.03762>], (02.08. 2023). Accessed: 10.03.2024.
- [3] maxime, “hat is a transformer?.” <https://medium.com/inside-machine-learning/what-is-a-transformer-d07dd1fbec04>, (Jan 4, 2019). Accessed: 09.03.2024.
- [4] İşnet blog, “gan nedir.” <https://www.isnet.net.tr/blogicerik/generative-ai-nedir/>, (Jan 4, 2019). Accessed: 09.03.2024.
- [5] A. Blattmann, “Stable video diffusion: Scaling latent video diffusion models to large datasets.” [<https://arxiv.org/abs/2311.15127>], (nov 25, 2023). Accessed:13.03.2024.
- [6] gensim, “Read the docs: documentation simplified.” <https://docs.readthedocs.io/>. Accessed: 13.05.2024.
- [7] G. Yakar and M. KINIK, “Yapay zekâ ile üretilen görsel sanatlar eserlerinde fikrî mülkiyet,” *Art-e Sanat Dergisi*, vol. 13, no. 26, pp. 490–516, 2020.
- [8] can sungur, “Stable diffusion ’ı anlamak: Negatif pozitif prompt, neural network.” [<https://www.youtube.com/watch?v=9YPoQFxbTqk>], (15.03, 2023). Accessed:02.11.2020.
- [9] C. ai, “models.” [<https://civitai.com/models/86963/landscapesupermix>], (15.03, 2024). Accessed:02.11.2020.
- [10] AUTOMATIC1111, “models.” [<https://github.com/AUTOMATIC1111>], (15.03, 2024). Accessed:02.11.2020.
- [11] E. Demir, “Transformer’ı anlamak — bölüm 1.” [https://demegire.medium.com/transformer\(03.03, 2024\)](https://demegire.medium.com/transformer(03.03, 2024)). Accessed:02.11.2020.
- [12] unsplash, “The most powerful photo engine in the world.” <https://unsplash.com/developers>. Accessed: 27.05.2024.
- [13] A. ROUGETET, “Landscape pictures.” [<https://www.kaggle.com/datasets/arnaud58/landscape-pictures>], (nov 15, 2020). Accessed:10.03.2023.

- [14] M. KALKENINGS, “Litter on forest floor.” [<https://www.kaggle.com/datasets/milankalkenings/litter-on-forest-floor>], (15.03, 2023). Accessed:10.03.2023.
- [15] J. BHATHENA, “Weather image recognition.” [<https://www.kaggle.com/datasets/jehanbhathena/weather-dataset>], (15.03, 2023). Accessed:10.03.2022.
- [16] X. Wang, Y. Huang, L. Li, H. Dai, F. Song, and K. Chen, “Assessment of performance of the gail model for predicting breast cancer risk: a systematic review and meta-analysis with trial sequential analysis,” *Breast Cancer Research*, vol. 20, pp. 1–19, 2018.