



Metinden görsel oluşturma projesi

Mustafa Erdoğan

14.03.2024

1 Giriş

Bu rapor, "Metinden görsel üretme" adlı projenin detaylarını içermektedir. Projenin temel amacı, metinden görsel üretebilen bir yapay zeka modeli geliştirebilmektir. Hedef, kullanıcı tarafından oluşturulan doğa tasvirini derin öğrenme yöntemleri kullanarak özgün ve yaratıcı doğa görselleri oluşturabilen bir sistem oluşturmaktır.

2 Literatür Araştırması

Kullanılan ve Benzer çalışmalar

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.[1]

2.1.1 spaCy nedir?

spaCy, Python programlama dilinde kullanılan güçlü bir açık kaynaklı doğal dil işleme (NLP) kütüphanesidir. spaCy, nlp kullanarak bilgisayarların insan dilini anlamasını ve işleyebilmesini sağlayan bir yapay zeka alanıdır. spaCy, metinleri analiz etmek ve anlam çıkarmak için çeşitli teknikler kullanır.[2]

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.[3]

2.2.1 Transformatörler nedir

'Attention Is All You Need ' başlıklı makale transformatörleri tanıtıyor. Başlığın da belirttiği gibi, daha önce gördüğümüz dikkat mekanizmasını kullanıyor.[4] 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[5]

2.3 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[6]

2.4 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. [7]

2.5 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.

3 Metodoloji

Kullanılan teknolojiler

3.1 Stable Video Diffusion nedir?

Stable Video Diffusion (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. [8]

3.2 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.[8]

3.3 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.

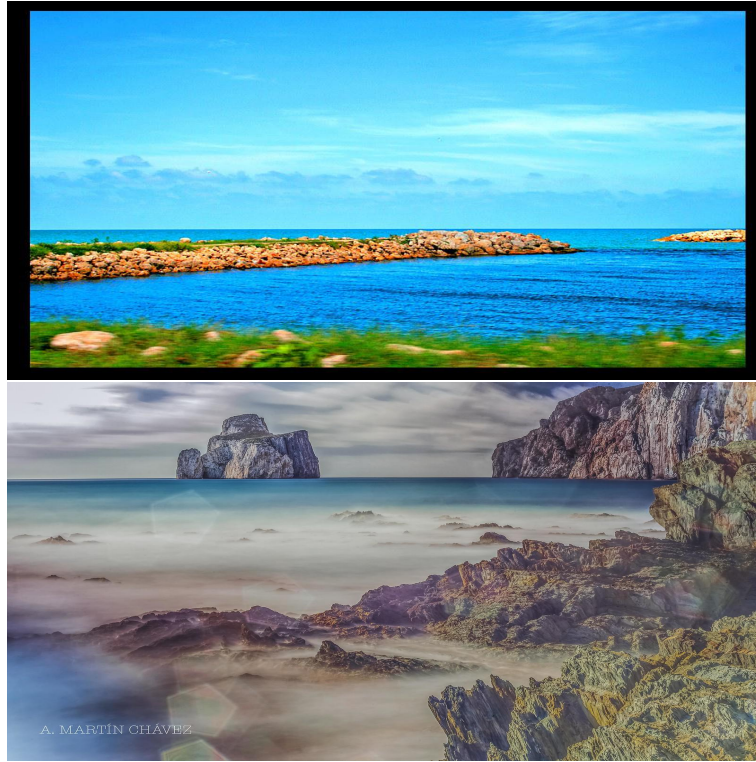
4 Veri Tabanı ve Veriler

Veri kümesinin içeriği:Doğa görselleri

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

Veri kaynağı:[9],[10],[11]

Veri örnekleri:



5 Beklenen Sonular

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 karşılaştırılacaktır. Bu sayede, Stable Video Diffusion modelinin güçlü ve zayıf yönleri belirlenerek, bu modele benzer bir modelin geliştirilmesine zemin hazırlanacaktır. Proje kapsamında, modelin metin açıklamalarını doğru şekilde yorumlayabilmesi için gerekli veri setleri ve eğitim yöntemleri belirlenecek, farklı manzara türlerini ve farklı metin formatlarını desteklemesi için gerekli geliştirmeler yapılacaktır. Projenin tamamlanmasıyla birlikte, kullanıcının girdiği basit metinden manzara resmi oluşturabilen bir model inşa edilmiş olacak ve bu modelin farklı metin açıklamalarına ve parametrelere karşı test edilmesi, modelin oluşturduğu resimlerin kalitesinin ve çeşitliliğinin değerlendirilmesi ve modelin sınırlamalarının belirlenmesi gibi işlemler tamamlanacaktır. Son olarak, projenin kapsamını, kullanılan yöntemleri, elde edilen sonuçları ve gelecekteki çalışmaları içeren bir rapor hazırlanacaktır. [12]

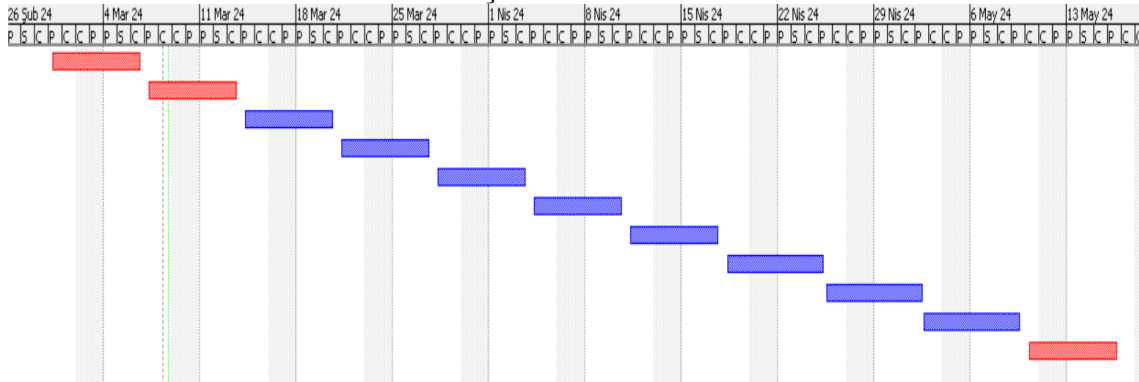


Powered by DALL·E 3

GANTT Chart iş paketlerinin tanımı

	Ad	Süre	Başlat	Bitirme
1	Proje belirlenmesi ve ders hakkında araştırma	5 günler	29.02.2024 08:00	06.03.2024 17:00
2	Proje kapsamını ve hedeflerini netleştirmek.	5 günler	07.03.2024 08:00	13.03.2024 17:00
3	Stable video diffusion araştırma ve sunum	5 günler	14.03.2024 08:00	20.03.2024 17:00
4	Transformer araştırması ve sunum	5 günler	21.03.2024 08:00	27.03.2024 17:00
5	Metin işleme ve görsel oluşturma modellerini eğitmek için veri hazırlama.	5 günler	28.03.2024 08:00	03.04.2024 17:00
6	Metin işleme ve görsel oluşturma modellerini eğitmek.	5 günler	04.04.2024 08:00	10.04.2024 17:00
7	Modellerin performansını değerlendirmek	5 günler	11.04.2024 08:00	17.04.2024 17:00
8	Kullanıcı arayüzü ve görselleştirme geliştirme.	5 günler	18.04.2024 09:00	25.04.2024 09:00
9	Projenin ilk prototipini tamamlamak.	5 günler	25.04.2024 14:00	02.05.2024 14:00
10	Kullanıcı geri bildirimlerine göre prototipi geliştirmek	5 günler	02.05.2024 16:00	09.05.2024 16:00
11	Projenin son raporunu ve sunumunu hazırlamak.	4 günler	10.05.2024 08:00	15.05.2024 17:00

Şekil 1



Şekil 2

Şekil 1 ve şekil 2’de görebileceği üzere iş akış planı gösterilmektedir.

Kaynaça

- [1] A. M. Hafiz.
- [2] spacy.
- [3] 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.
- [4] A. Vaswani, “Attention is all you nee.”
- [5] maxime, “hat is a transformer?.” <https://medium.com/inside-machine-learning/what-is-a-transformer-d07dd1fbec04>, (Jan 4, 2019). Accessed: 09.03.2024.
- [6] İşnet blog, “gan nedir.” <https://www.isnet.net.tr/blogicerik/generative-ai-nedir/>, (Jan 4, 2019). Accessed: 09.03.2024.
- [7] A. Blattmann.
- [8] 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.
- [9] A. ROUGETET.
- [10] M. KALKENINGS.
- [11] J. BHATHENA.
- [12] 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.