logo, yazı tipi, simge, sembol, grafik içeren bir resim

Açıklama otomatik olarak oluşturuldu

Kedi, Köpek, Vahşi Hayvan Sınıflandırma

Hüseyin Mete Çalışkan 190101048,

Lütfi Yiğit Sağlam 190101090,

Sude Özdemir 190101032

1. Giriş
   1. **Projenin Amacı**

Bu proje, bilgisayarlı görüntü ve makine öğrenimi tekniklerini kullanarak kedi, köpek ve vahşi hayvan türlerini sınıflandırmayı amaçlamaktadır. Günümüzde, biyolojik çeşitlilik izleme ve koruma çabaları için doğru ve etkili hayvan türü tanımlama yöntemleri büyük önem taşımaktadır. Ancak, manuel olarak yapılan bu tanımlamalar zaman alıcı ve maliyetlidir. Bu nedenle, bu proje, görüntü tanıma algoritmaları ve makine öğrenimi modelleri kullanarak, çeşitli hayvan türlerini otomatik olarak sınıflandırmak için bir sistem geliştirmeyi hedeflemektedir.

Bu projenin sonuçları, biyolojik çeşitlilik araştırmacılarına, doğal yaşamı koruma uzmanlarına ve diğer ilgili paydaşlara, hızlı ve doğru bir şekilde farklı hayvan türlerini tanımlama ve izleme konusunda yardımcı olabilecek bir araç sunacaktır. Ayrıca, bu projenin başarısı, benzer doğal yaşam projeleri için bir referans noktası olabilir ve benzer çalışmaların ilerlemesine katkıda bulunabilir.

**1.2 Hedefler**

Doğru Sınıflandırma: Farklı hayvan türlerini doğru bir şekilde tanımlayan bir model geliştirmek.

Yüksek Performans: Modelin, veri seti üzerinde yüksek doğruluk oranı göstermesini sağlamak.

Algoritmaların Uygulanması: Öğrenilen örüntü tanıma algoritmalarını başarılı bir şekilde uygulamak.

Sonuçların Değerlendirilmesi: Elde edilen sonuçların analiz edilmesi ve gelecek çalışmalara yönelik önerilerin sunulması.

Dökümantasyon ve Sunum: Projenin eksiksiz bir şekilde dokümante edilmesi ve uygun bir sunumun hazırlanması.

1. Metodoloji
   1. **Araştırma Soruları**
2. Veri Toplama Süreci

Projede kullanılan veri setlerinin kaynakları ve özellikleri nelerdir?

2. Örüntü Tanıma Algoritmalarının Seçimi

Projede hangi örüntü tanıma algoritmaları kullanılmıştır?

Seçilen algoritmaların projeye uygunluğu nasıl değerlendirilmiştir?

3. Model Eğitimi ve Değerlendirme

Kullanılan örüntü tanıma modellerinin eğitim süreci nasıl gerçekleştirilmiştir?

Modellerin performansı nasıl değerlendirilmiştir? Hangi metrikler kullanılmıştır?

4. Sonuçların Analizi ve Tartışması

Projede elde edilen sonuçlar neyi göstermektedir? Karşılaşılan zorluklar ve başarılar nelerdir? Elde edilen bulguların pratik uygulamalara ve gelecekteki araştırmalara etkisi nedir?

**2.2 Veri Toplama**

Veri toplama süreci, Kaggle platformu üzerinden erişilebilen "Animal Faces" adlı veri seti kullanılarak gerçekleştirilmiştir. Veri seti, Animal Faces-HQ (AFHQ) olarak da bilinen 4217 adet yüksek kaliteli görüntüden oluşmaktadır ve 512×512 çözünürlüğe sahiptir. Veri seti, üç farklı sınıf alanından oluşmaktadır. Her bir sınıf alanı, kedi, köpek ve vahşi yaşam gibi farklı hayvan türlerini içermektedir. Bu çoklu sınıf alanları ve her bir sınıf alanındaki çeşitli ırklara ait görüntüler, AFHQ'yu zorlu bir görüntüden görüntüye çeviri problemi haline getirmektedir. Bu veri seti, özellikle çeşitli sınıf alanları arasındaki farklılıklar ve çeşitli ırkların görüntüleri sayesinde, geniş bir veri yelpazesi sunmaktadır. Bu, özellikle görüntüden görüntüye çeviri algoritmaları için zorlu bir deneyim sunmaktadır ve çeşitli hayvan türlerini başarılı bir şekilde tanımlamak için modelin derinlemesine öğrenme yeteneğini test etmektedir.

Veri setimizi hazırlarken, öncelikle görüntülerin boyutlarını standartlaştırdık, böylece tüm görüntüler aynı boyuta yeniden boyutlandırıldı. Ardından, veri arttırma teknikleri uygulayarak görüntü çeşitliliğini artırdık. Bu adımlar arasında, görüntüleri döndürme, yansıtma, kaydırma ve zoom gibi dönüşümler kullanıldı.

**A screen shot of a computer

Description automatically generated**

**A screenshot of a computer program

Description automatically generated**

**2.3 Veri Analizi**

Verilerin bilgisayara aktarılması ve analizi, Python programlama dili kullanılarak gerçekleştirilmiştir. İlk olarak, Kaggle platformundan indirilen veri seti, Python kütüphaneleri aracılığıyla bir veri çerçevesine yüklenmiştir. Veri çerçevesi üzerinde yapılan ön işleme adımları, özellikle resim dosyalarının yüklenmesi ve boyutlandırılması gibi işlemleri içermiştir. Ardından, veri seti üzerinde örüntü tanıma algoritmalarını uygulamak için TensorFlow ve Numpy gibi makine öğrenimi kütüphaneleri kullanılmıştır. Model eğitimi için veri seti, eğitim ve test verisi olarak ayrılmış ve doğruluk değerlerini hesaplamak için çapraz doğrulama yöntemleri kullanılmıştır. Sonuçlar, grafikler ve performans metrikleri aracılığıyla analiz edilmiş ve yorumlanmıştır. Bu süreç, projenin temel analiz aşamasını oluşturmuş ve verilerin bilgisayar ortamında nasıl işlendiğini göstermiştir.

**2.4 Sınırlamalar**

Çalışmanın tasarımı bazı sınırlamalara tabidir. Öncelikle, veri seti çeşitli internet kaynaklarından derlenmiş olup, bu kaynakların doğruluğu ve güvenilirliği kontrol edilememiştir. Bu durum, veri setindeki etiketlenmiş sınıfların tam doğruluğunu garanti etmemektedir. Ayrıca, veri setinin boyutu sınırlı olup, özellikle bazı nadir sınıfların temsil edilme oranı düşüktür. Bu durum, modelin nadir sınıfları tanıma yeteneğini etkileyebilir ve genelleme yeteneğini kısıtlayabilir. Dahası, projede kullanılan örüntü tanıma algoritmaları, belirli sınırlamalar ve varsayımlar altında çalışmaktadır ve bu da sonuçların genelleme gücünü etkileyebilir. Biz projemizde CNN, VGG16 ve MLP algoritmalarını kullandık ve algoritmalar için veri miktarı ve kalitesi, hesaplama gücü, overfitting, hiperparametre ayarı, veri seti dengesi, özellik mühendisliği gibi sınırlamalar olabilir. Bu sınırlamalar, projenin sonuçlarının genel geçerliliğini sınırlayabilir ve dikkate alınması gereken önemli faktörlerdir.

**2.5 Araştırma Tasarımı**

Projenin amacı, "Hayvan Sınıflandırma" adlı veri seti üzerinde çeşitli örüntü tanıma algoritmalarını uygulayarak farklı hayvan türlerini doğru bir şekilde sınıflandırmaktır. Bu amaç doğrultusunda, aşağıdaki araştırma sorularını yanıtlamak için belirli bir metodoloji izlenecektir:

Araştırma Sorusu: Hayvan türlerini doğru bir şekilde sınıflandırmak için hangi örüntü tanıma algoritmaları kullanılabilir?

Veri Seti: Kaggle platformundan temin edilen hayvan görüntülerini içeren bir veri seti kullanılacaktır. Veri seti, köpekler, kediler ve vahşi hayvanlar gibi farklı hayvan türlerinin görüntülerinden oluşmaktadır.

Metodoloji:

Model eğitiminde CNN, VGG16 ve MLP kullanılacaktır.

Model eğitimi ve doğrulamasında test ve “train” veri setleri kullanılacaktır.

Eğitilen modelin performansı, doğruluk, hassasiyet ve F1 skoru gibi metriklerle değerlendirilecektir.

Sonuçlar: Elde edilen sonuçlar, kullanılan örüntü tanıma algoritmalarının hayvan türlerini doğru bir şekilde sınıflandırma yeteneğini gösterecektir. Bu sonuçlar, hayvan türü tespiti gibi uygulamalarda kullanılan algoritmaların etkinliğini değerlendirmeye yardımcı olacaktır.

Öneriler: Araştırma sonuçlarına dayanarak, daha geniş kapsamlı bir veri seti kullanarak ve modelin karmaşıklığını artırarak algoritmanın performansının daha da iyileştirilebileceği önerilebilir.

3 Sonuçlar

**3.1 Araştırma Soruları**

Hayvan türlerini doğru bir şekilde sınıflandırmak için hangi örüntü tanıma algoritmaları kullanılmıştır?

Biz bu proje için CNN, VGG16, MLP algoritmasını kullandık.

Modelin eğitilmesi ve doğrulanması için hangi metrikler ve yöntemler kullanılacaktır? Modelin performansı hangi kriterlere göre değerlendirilecek ve raporlanacaktır?

Accuracy, confusion matrix ve scikit-learn kütüphanesinin metrics modülü altında bulunan classification report fonksiyonunu kullandık.

Projede hangi tür hayvanların tanınması hedeflenmektedir ve bu hayvanların veri seti ne kadar çeşitlidir?

Kedi, köpek ve vahşi hayvan türlerini ayırt etmek için bir örüntü tanıma modeli geliştirdik. Train veri setimizde toplam 4217 fotoğraf örneği bulunmaktadır. Test kısmı içinse 1430 fotoğraf bulunmaktadır.

Hangi tür veri setleri kullanılacak ve bu veri setleri ne kadar çeşitli olacak?

Etiketlenmiş veri seti kullandık ve veri setimiz test, train olmak üzere iki parçadır.

Projede kullanılan veri setleri ve modellerin etiketlenmesi ve doğruluğu nasıl sağlanacaktır?

Projemizde etiketlenmiş veri seti kullandık.

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**memeli, tilki, kızıl tilki, kum tilkisi içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**A screenshot of a computer program

Description automatically generated**

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer screen

Description automatically generated



metin, ekran görüntüsü, diyagram, dikdörtgen içeren bir resim

Açıklama otomatik olarak oluşturuldu

A screenshot of a computer

Description automatically generated



|  |  |
| --- | --- |
| Model adı | Accuracy |
| CNN | 0.90 |
| VGG16 | 0.87 |
| MLP | 0.73 |

4 Tartışma

Sonuçlarımızda dikkat çeken bazı önemli noktalar bulunmaktadır. Özellikle, örüntü tanıma algoritmalarının doğruluk, hassasiyet ve F1 skoru gibi metriklerle ölçülen performansı üzerinde durmak önemlidir. Modelimizin yüksek doğruluk oranları elde ettiği görülmektedir, ancak beklenmedik bir şekilde bazı sınıfların tanınmasında zayıf sonuçlar elde edilmiştir. Bu sonuçlar, modelin belirli hayvan türlerini diğerlerine göre daha az başarıyla tanıma eğiliminde olduğunu göstermektedir.

Bu sonuçların birkaç olası nedeni vardır. Birincisi, veri setindeki dengesizlik modelin performansını etkileyebilir. Özellikle, daha az temsil edilen sınıfların tanınmasında zorluklar yaşanabilir. Ayrıca, modelin özellik seçimi ve eğitimi sürecinde kullanılan parametrelerin optimal olmayabileceği de göz önünde bulundurulmalıdır. Belirli hayvan türlerine ait görüntülerin karmaşıklığı ve çeşitliliği de sonuçları etkileyebilir.

Bu sonuçların programın sonuçlarına ve etkisine olan etkileri önemlidir. Yüksek doğruluk oranlarına rağmen, modelin belirli hayvan türlerini başarılı bir şekilde tanıyamaması, uygulamanın gerçek dünya kullanımında güvenilirliğini sorgulatabilir. Bu nedenle, modelin daha geniş bir veri setiyle eğitilmesi ve özellik seçimi sürecinin iyileştirilmesi programın etkisini artırabilir.

Çalışmanın bazı sınırlamaları bulunmaktadır. Özellikle, kullanılan veri setinin sınırlı olması ve dengesiz dağılımı sonuçları etkileyebilir. Ayrıca, modelin eğitim sürecinde kullanılan parametrelerin belirlenmesindeki belirsizlikler de sonuçları etkileyebilir. Bu sınırlamaların bilinmesi, çalışmanın yorumlanması ve sonuçların doğru şekilde değerlendirilmesi açısından önemlidir.

# 5 Öneriler ve Sonuç

Veri Seti Çeşitliliği Artırılmalı: Modelin genelleme yeteneğini artırmak için, kullanılan veri setinin çeşitliliği artırılmalıdır. Farklı çevresel koşullar, ışıklandırma ve pozlar içeren daha geniş bir veri seti, modelin çeşitli koşullarda daha iyi performans göstermesini sağlayabilir.

Özellik Seçimi ve Model Parametreleri Optimize Edilmeli: Modelin doğruluğunu artırmak için, özellik seçimi süreci ve model parametreleri daha dikkatli bir şekilde optimize edilmelidir. Daha etkili özelliklerin seçilmesi ve modelin daha iyi performans göstermesini sağlayacak optimal parametrelerin belirlenmesi, genel performansı önemli ölçüde artırabilir.

Dengeli Veri Seti Kullanılmalı: Modelin eğitiminde kullanılan veri seti dengeli olmalıdır. Her bir hayvan türünden yeterli sayıda örneğin bulunduğu bir dengeli veri seti, modelin farklı sınıfları daha dengeli bir şekilde tanımasına yardımcı olabilir ve yanlılığı önleyebilir.

6 Referanslar

<https://www.kaggle.com/code/ulfbereth/hayvan-siniflandirma1/input>

<https://chatgpt.com/>

<https://www.geeksforgeeks.org/pattern-recognition-introduction/>

<https://translate.google.com/?sl=en&tl=tr&op=translate>

<https://www.kaggle.com/>

<https://colab.research.google.com/>

[https://www.nobledesktop.com/classes-near-me/blog/pandas-vs-numpy-for-data-analytics - :~:text=Pandas is most commonly used,makes N-dimensional homogeneous objects.](https://www.nobledesktop.com/classes-near-me/blog/pandas-vs-numpy-for-data-analytics%20-%20:~:text=Pandas%20is%20most%20commonly%20used,makes%20N-dimensional%20homogeneous%20objects.)

<https://www.sciencedirect.com/science/article/pii/S1090023306002206>

<http://advances.utc.sk/index.php/AEEE/article/view/2202/1271>