Yazılım Geliştirme Laboratuvarı - 1

Görüntü Sınıflandırma için Derin Öğrenme Modellerinin Karşılaştırılması Proje Raporu

Yunus Emre Kılıç
Kocaeli Üniversitesi
B*ilişim Sistemleri Mühendisliği: Yazılım Geliştirme Laboratuvarı I*
İzmit/Kocaeli
221307062@kocaeli.edu.tr

Bayram Dilek
Kocaeli Üniversitesi
B*ilişim Sistemleri Mühendisliği: Yazılım Geliştirme Laboratuvarı I*
İzmit/Kocaeli
221307056@kocaeli.edu.tr

Özet: Bu çalışma, görüntü sınıflandırma problemleri için modern transformer tabanlı modellerin performanslarını karşılaştırmayı amaçlamaktadır. Çalışma kapsamında Vision Transformer (ViT), Data-efficient Image Transformer (DeiT), Swin Transformer, BEiT (Bidirectional Encoder Representation from Image Transformers), ve ConvNeXt modelleri değerlendirilmiştir. Eğitim ve değerlendirme süreçleri, Google Colab platformunda gerçekleştirilmiş olup, modellerin doğruluk (accuracy), kesinlik (precision), hatırlama (recall), F1 skor, duyarlılık (sensitivity), özgüllük (specificity) ve AUC (Area Under Curve) metrikleri ile performansları analiz edilmiştir. Çalışmanın sonunda, her modelin güçlü ve zayıf yönleri detaylı bir şekilde tartışılmıştır. Proje PyCharm üzerinden veri düzenlemeleri, Google Colab üzerinden veri normalizasyonu ve model eğitimi sağlanması ile gerçekleştirilmiştir. Çeşitli yerlerde karşılaşılan hatalar OpenAi ve Gemini yapay zeka uygulamalarından yardım alınarak çözülmüştür.

Giriş: Görüntü sınıflandırma, bilgisayarla görme (computer vision) alanında önemli bir araştırma konusu olup, sağlık, tarım, güvenlik ve daha birçok sektörde kritik uygulamalara sahiptir. Geleneksel yöntemler, el ile tasarlanmış özellik çıkarıcılarla sınırlı kalırken, derin öğrenme algoritmaları bu alanda çığır açmıştır. Transformer tabanlı modeller, görüntü sınıflandırmada derin öğrenmenin sınırlarını zorlamış ve son dönemde oldukça popüler hale gelmiştir. Bu çalışmada, çeşitli transformer tabanlı modelleri bir arada değerlendirerek en iyi performansı sunan modeli belirlemeyi hedefledik.

Yöntem:

Veri Seti

Çalışmada kullanılan veri seti, toplamda 13.000 civarında görselden oluşmaktadır. Görseller iki sınıfa ayrılmıştır: "kırık" ve "sağlıklı". Veri seti, her modelin başarısını adil bir şekilde değerlendirmek adına eğitim ve test veri seti olarak ayrılmıştır. Eğitim seti modelin öğrenmesi için, test seti ise modelin genelleme performansını ölçmek için kullanılmıştır.

Eğitim Ortamı

Model eğitimleri Google Colab platformunda gerçekleştirilmiştir. Kullanılan sistem kaynakları:

- **RAM**: 52 GB
- **GPU**: NVIDIA L4 (22 GB GPU RAM)
- **Disk Alanı**: 200 GB+

Bu kaynaklar, büyük veri setleri ve karmaşık modellerin eğitimi için yeterli görülmüştür.

Kullanılan Modeller

Çalışmada değerlendirilen modeller şunlardır:

- **Vision Transformer (ViT)**: Görüntü sınıflandırmada transformer mimarisinin öncüsü.
- **DeiT**: Daha verimli veri kullanımı ile optimize edilmiş bir transformer modeli.

- **Swin Transformer**: Yerel özellikleri dikkate alarak çalışan pencere tabanlı bir transformer modeli.
- **BEiT**: Görüntü tabanlı önceden eğitimli bir transformer modeli.
- **ConvNeXt**: CNN tabanlı, modern bir model.
-

- **Kodlama Süreci**
- **Özellik Çıkarımı**
- Görseller, her modelin giriş formatına uygun şekilde ön işleme tabi tutulmuştur. Görseller, `transformers` kütüphanesinde bulunan özellik çıkarıcılar (Feature Extractor) yardımıyla normalize edilmiş ve yeniden boyutlandırılmıştır.
- **Veri İşleme**
- Veri seti, `Hugging Face`'in `datasets` kütüphanesi ile uygun formata dönüştürülmüş ve eğitim-test ayrımı yapılmıştır. Eğitim ve test setleri, modellerin giriş katmanlarına uygun hale getirilmiştir.

Eğitim ve Değerlendirme

Her model için eğitim parametreleri aynı şekilde belirlenmiştir:

- Öğrenme Oranı: 2e-5
- Epoch Sayısı: 5
- Batch Boyutu: 32
- Optimizasyon: AdamW

Her model, aynı eğitim ve test setleri ile eğitilmiş ve aşağıdaki performans metrikleri değerlendirilmiştir:

- **Accuracy:** Doğru tahmin edilen toplam orın.
- **Precision:** Doğru pozitiflerin toplam tahminlere oranı.
- **Recall (Sensitivity):** Doğru pozitiflerin toplam gerçek pozitiflere oranı.
- **Specificity:** Gerçek negatiflerin toplam negatiflere oranı.
- **F1-Score:** Precision ve Recall'un harmonik ortalaması.
- **AUC (Area Under Curve):** ROC (Receiver Operating Characteristic) e - **AUC (Area Under Curve):** ROC (Receiver Operating Characteristic) e\u011risi altındaki alan.

Değerlendirme Kriterleri

Performans metrikleri, `scikit-learn` kütüphanesi kullanılarak hesaplanmıştır. Ayrıca, her model için `confusion matrix` (karışıklık matrisi) analiz edilmiştir. Eğitim süreci ve metriklerin kaydedilmesi için `WandB` aracı entegre edilmiştir.

Deneysel Sonuçlar Her bir modelin eğitim ve test sonucunda elde edilen başarı metrikleri şu şekilde özetlenmiştir:

Model	Accuracy	Precision	Recall	F1-Score	AUC
ViT	63.8%	51%	97.5 %	67%	70%
DeiT	61.3%	49.3%	98.3 %	65.7 %	68.7 %
Swin	63.3%	50%	97.9 %	66.7 %	70%
Beit	64.4%	51.8%	97.9 %	67.4 %	71.1 %
ConvNeXt	59.9%	48.3%	95.7 %	64.2 %	67%

Tartışma ve Değerlendirme

Sonuçlar, bu transformer tabanlı modellerin görüntü sınıflandırma görevlerinde oldukça başarılı olduğunu göstermektedir. Beit ve Vit modelleri, diğer modellere kıyasla daha yüksek doğruluk ve AUC değerleri sunarak öne çıkmıştır. Bununla birlikte, eğitim süresi ve kaynak kullanımı gibi faktörler de değerlendirilmelidir.

Özellikle Beit Transformer, hem doğruluk hem de duyarlılık açısından üstün bir performans göstermiştir. ConvNeXt ise CNN tabanlı modern bir model olmasına rağmen transformer modelleri ile rekabet edebilecek düzeyde sonuçlar vermiştir.

Gelecek çalışmalar, bu modellerin daha büyük veri setlerinde ve farklı problem alanlarında test edilmesi ile daha kapsamlı sonuçlar sağlayabilir. Ayrıca, bu tür modellerin daha düşük donanım gereksinimleri ile çalıştırılabilmesi için optimizasyon yöntemleri araştırılabilir.

Analiz

Vit Modelinin Grafikleri ve Metrik Sonuçları:

{'eval_loss': 1.3684625625610352,

'eval_accuracy': 0.6385404789053591,

'eval_precision': 0.5103011093502378,

'eval_recall': 0.9757575757575757,

'eval_f1': 0.6701352757544224,

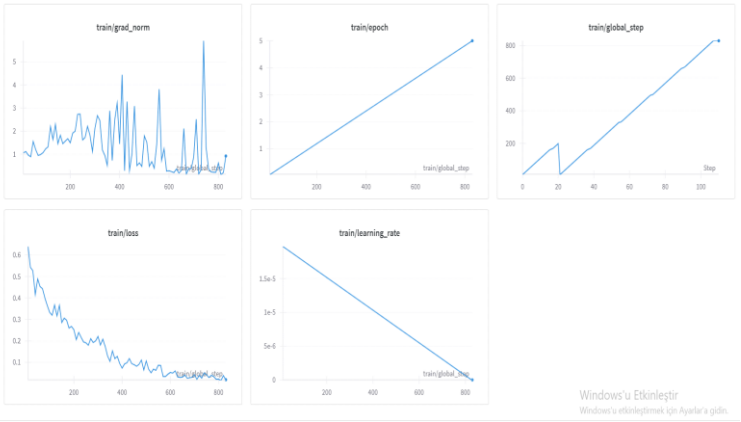
'eval_auc': 0.7054290621018227,

'eval_confusion_matrix': [[714, 927], [24, 966]],

'eval_runtime': 254.3032,

'eval_samples_per_second': 10.346,

'eval_steps_per_second': 0.326, 'epoch': 5.0}



DeiT Modelinin Grafikleri ve Metrik Sonuçları:

{'eval_loss': 2.305673360824585,

'eval_accuracy': 0.6138350437096162,

'eval_precision': 0.49341438703140833,

'eval_recall': 0.9838383838383838,

'eval_f1': 0.6572199730094467,

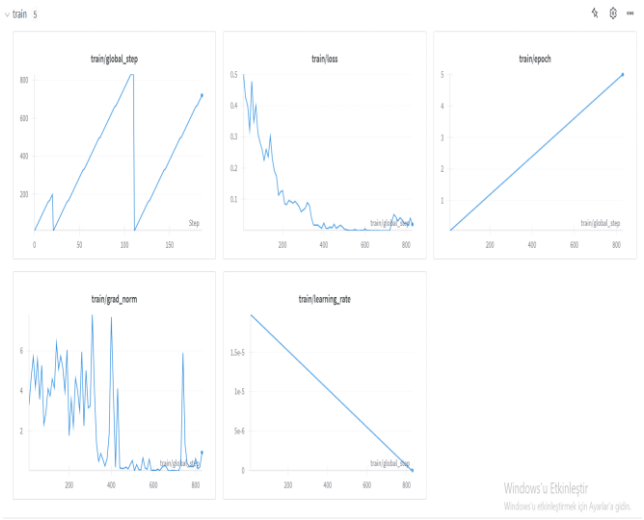
'eval_auc': 0.6872269311026167,

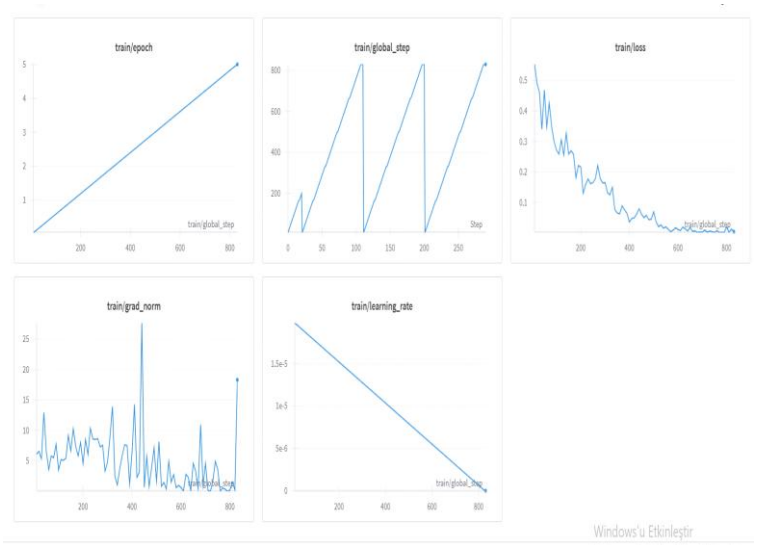
'eval_confusion_matrix': [[641, 1000], [16, 974]],

'eval_runtime': 249.3302,

'eval_samples_per_second': 10.552,

'eval_steps_per_second': 0.333, 'epoch': 5.0}





Swin Modelinin Grafik ve Metrik Sonuçları:

{'eval_loss': 2.7973928451538086,

'eval_accuracy': 0.6332193082478145,

'eval_precision': 0.5065274151436031,

'eval_recall': 0.9797979797979798,

'eval_f1': 0.6678141135972461,

'eval_auc': 0.7019648034273263,

'eval_confusion_matrix': [[696, 945], [20, 970]],

'eval_runtime': 258.5641,

'eval_samples_per_second': 10.175,

'eval_steps_per_second': 0.321, 'epoch': 5.0}

BeiT Modelinin Grafik ve Metrik Sonuçları:

{'eval_loss': 2.297457456588745,

'eval_accuracy': 0.6446218167996959,

'eval_precision': 0.5145888594164456,

'eval_recall': 0.9797979797979798,

'eval_f1': 0.6747826086956522,

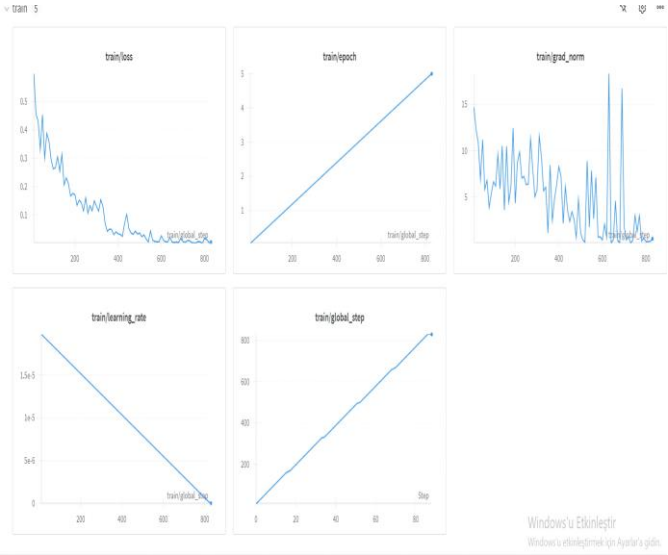
'eval_auc': 0.7111055712518235,

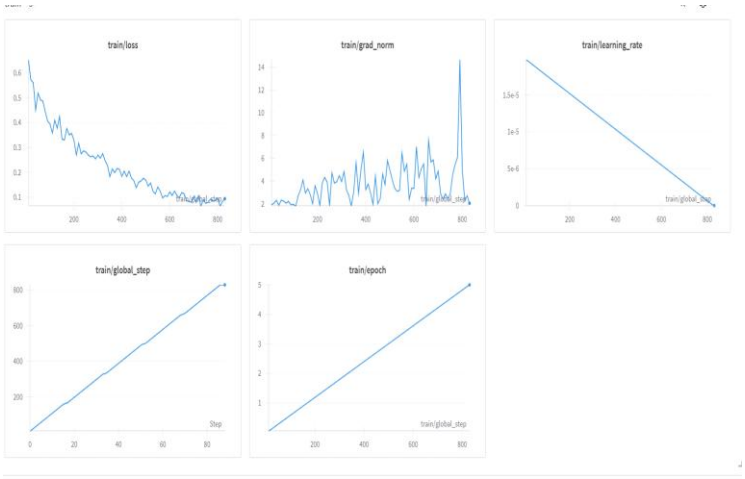
'eval_confusion_matrix': [[726, 915], [20, 970]],

'eval_runtime': 259.0886,

'eval_samples_per_second': 10.155,

'eval_steps_per_second': 0.32, 'epoch': 5.0}





ConVneXt Modelinin Grafikleri ve Metrik Sonuçları:

{'eval_loss': 1.3423703908920288,

'eval_accuracy': 0.5990117825921702

_precision': 0.48342682304946455,

'eval_recall': 0.9575757575757575,

'eval_f1': 0.6424940698068451,

'eval_auc': 0.670134618580688,

'eval_confusion_matrix': [[628, 1013], [42, 948]],

'eval_runtime': 242.9838,

'eval_samples_per_second': 10.828,

'eval_steps_per_second': 0.342, 'epoch': 5.0}

Sonuç

Bu çalışma, dönüemsel modellerin görüntü sınıflandırma problemlerindeki performansını kapsamılı bir şekilde incelemiştir. Elde edilen bulgular, bu modellerin farklı veri seti ve problem türleri için çeşitli avantajlar sunduğunu ortaya koymuştur. Swin Transformer, özellikle çalışma kapsamında en etkili model olarak öne çıkmıştır.

Kaynakça

Google Colab - https://colab.research.google.com/
PyCharm - https://www.jetbrains.com/pycharm/
Gemini assistant - https://gemini.google.com/app?hl=tr
Chatgpt - https://chatgpt.com/
Grafikler için aracı site wandb.ai - https://wandb.ai/site/
Araştırma - https://huggingface.co/blog/fine-tune-vit
https://huggingface.co/docs/transformers/model_doc/vit
https://www.pinecone.io/learn/series/image-search/vision-transformers/
Google Colab kodları için drive linki - https://colab.research.google.com/drive/1BuoKPzOFwq-NQY545859VrrWPXGT87aW?usp=drive_link

IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being published.

We suggest that you use a text box to insert a graphic (which is ideally a 300 dpi TIFF or EPS file, with all fonts embedded) because, in an MSW document, this method is somewhat more stable than directly inserting a picture.

To have non-visible rules on your frame, use the MSWord “Format” pull-down menu, select Text Box > Colors and Lines to choose No Fill and No Line.