

**Swin Tabanlı Dönüştürülmüş Görüntülerin Sınıflandırılması**Serdar ÇİFTÇİ<sup>1,\*</sup><sup>1</sup>Harran Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, 63050, Haliliye/Şanlıurfa**Öz**

Görüntü sınıflandırma bilgisayarlı görü alanındaki temel çalışmalardan biridir. Görüntü çözünürlüğü ve görüntünün keskinliği sınıflandırma performansını önemli ölçüde etkileyen faktörlerdendir. Bu çalışmada görüntülerin çözünürlüğünün ve keskinliğinin Swin tabanlı dönüştürücü olan Swin2SR algoritması kullanılarak artırılmasıyla görüntü sınıflandırma performansının değişimi incelenmiştir. Sınıflandırma için transfer öğrenme mimarilerinden biri olan ResNet18 modeli kullanılmıştır. CIFAR10 veri kümesi üzerinde 50 epok için yapılan deneyler sonucunda Swin2SR algoritmasının görüntülerin çözünürlüğünü ve netliğini artırarak sınıflandırma doğruluğunu %85'ten %87'ye çıkardığı gözlemlenmiştir..

**Makale Bilgisi**

Başvuru: 27/07/2023

Yayın: 31/08/2023

**Anahtar Kelimeler***Görüntü Sınıflandırma*  
*Swin2SR Dönüştürme*  
*Transfer Öğrenme***Keywords***Image Classification*  
*Swin2SR Transformer*  
*Transfer Learning***Classification of Swin-based Transformed Images****Abstract**

Image classification is one of the fundamental studies in computer vision. The resolution and sharpness of images are essential factors that affect classification performance. This study investigates how image classification performance changes by increasing image resolution and sharpness using the Swin2SR algorithm, a Swin-based transformer. The ResNet18 model, one of the transfer learning architectures, is used for image classification. The experiments conducted on the CIFAR10 dataset for 50 epochs have shown that the Swin2SR algorithm increases the classification accuracy from 85% to 87% by increasing the resolution and sharpness of the images.

**1. GİRİŞ (INTRODUCTION)**

Görüntü sınıflandırma, bilgisayarlı görü alanında sıkça çalışılmış konuların başında gelmektedir. Tarım, sağlık, medikal, savunma, eğlence ve daha birçok uygulama alanlarında temel işlem olarak kullanılan görüntü sınıflandırma uzunca bir süre makine öğrenmesi temelli yöntemler ile gerçekleştirilirken, 2012 yılında ImageNet yarışmasında derin öğrenme tabanlı AlexNet [1] mimarisinin başarısından sonra takip eden çalışmaların çoğunluğu evrimsel sinir ağları temelli olarak geliştirilmiştir [2-6].

Derin öğrenme mimarilerinin yüksek performans sergilemesi, eğitimleri sırasında büyük veri kümelerini kullanmasına ve eğitim hesaplamalarının yüksek başarımlı bilgisayarlar üzerinde gerçekleştirmesine bağlıdır. Gerek yüksek miktarda veri kümelerine erişimin her zaman mümkün olamaması gerekse de bu verilerin işlenebilmesinin yüksek maliyet oluşturması bu modellerin eğitimini zorlayabilmektedir. Yalnız, bu derin ağ mimarilerinin temel ilk katmanlarının ilkel verileri çıkardığı/tanıdığı görülmüştür. Her ne kadar ilk katman parametre değerleri farklı problem alanları için eğitilmiş olsa da başka problem alanlarında da benzer ilkel verileri başarıyla çıkardığı/tanıdığı gözlemlenmiştir [7]. Dolayısıyla ilk katmanların eğitilmeden doğrudan başka problem alanlarında da kullanılması yaygın olmuş ve bu yaklaşıma transfer öğrenme denmiştir. Bu çalışmada görüntü sınıflandırma için transfer öğrenme mimarilerinden biri olan ResNet18 [3] mimarisi kullanılmıştır.

Çözünürlük, gürültü, keskinlik ve kontrast gibi çeşitli faktörler görüntü sınıflandırma probleminin performansını etkilemektedir [8]. Görüntü çözünürlüğünün artırılması ile yapılan görüntü sınıflandırma

\*İletişim yazarı, e-mail: serdarçiftci@harran.edu.tr

çalışmalarında [9-11] başarımın arttığı gözlenmiş olup, bu makalede de çözünürlük ve keskinlik artırımının görüntü sınıflandırma üzerindeki etkisi incelenmiştir.

Görüntülerin düşük çözünürlükten yüksek çözünürlük haline getirilmesine süper-çözünürlük adı verilmektedir. Süper-çözünürlük yapılırken görüntü işleme [12], makine öğrenmesi [13] ve derin öğrenme tabanlı yaklaşımlar [14] kullanılmaktadır. Derin öğrenme tabanlı çalışmalardan evrişimsel [15], çekişmeli üretken ağlar [16] ve görü dönüştürücüler [17] süper-çözünürlük uygulamalarında yaygın olarak kullanılmakta ve başarılı performanslar sergilemektedir.

Dönüştürücüler ilk olarak doğal dil işleme alanında kullanılmış ve kendi dikkat mekanizması (self attention) sayesinde uzun sıralı bilgilerin korunumu başarılı bir şekilde mümkün olmuştur [18]. Dönüştürücülerin görüntü işleme alanında kullanımında [19]; görüntü bölütlere ayrılıp çok kafalı dikkat mekanizmaları sayesinde görüntü iyileştirme, bölütleme ve süper-çözünürlük gibi uygulamalarda başarılı sonuçlar ortaya konulmuştur [20]. Swin2SR [21] çalışması Swin tabanlı bir dönüştürücü [22] olup, süper-çözünürlük ve görüntü iyileştirmede başarılı sonuçlar üretmektedir. Bu makalede Swin2SR dönüştürücüsünün, görüntü çözünürlüğünü artırma ve görüntü keskinleştirme sağlamasıyla görüntü sınıflandırma üzerindeki etkisi incelenmiştir.

## 2. MATERYAL VE METOT (MATERIAL AND METHOD)

### 2.1 Swin2SR ile Süper-Çözünürlük ve Görüntü Keskinleştirme

Görü dönüştürücüler [19] görüntüyü bölütlere ayırarak işlem yapar ve görüntü iyileştirme, bölütleme ve süper-çözünürlük çalışmalarında başarılı sonuçlar verir. Swin dönüştürücü [22] algoritması ise görüntüyü hiyerarşik bir şekilde işlemektedir. Swin dönüştürücülerde kayan pencereler içerisinde kalan alanlar kendi dikkat (self-attention) mekanizmasına tabii tutulur. Bu çalışmada Swin dönüştürücüsünün bir türevi olan ve başarılı sonuçlar üreten Swin2SR modeli kullanılmıştır. Swin2SR algoritmasında katmanlar arası geçişte katman öncesi normalizasyon yerine katman sonrası normalizasyon uygulanmıştır. Böylelikle derin katmanlarda ortalama özellik varyans değişimi azalmış ve model eğitimi sırasında stabilite oluşmuştur. Swin2SR dönüştürmesi süper çözünürlük yapma, görüntü bozulmalarını giderme ve JPEG sıkıştırmasından kaynaklı bozulmaları gidermede başarılı performans sergilemiştir. Bu makalede süper çözünürlük için Swin2SR dönüştürmesinin çözünürlüğü 2 kat yükselten klasik (ClassicalSR\_X2) seçeneği kullanılmıştır. Model için yazarların github adresindeki kodlar<sup>1</sup> kullanılmıştır.

### 2.2 ResNet18 Mimarisi ile Görüntü Sınıflandırma

ResNet18 [3] mimarisi, artık bloklardan (residual blocks) oluşan bir derin öğrenme modelidir. Artık blokların özelliği, atlama bağlantılarının (skip connections) kullanılmasıyla girdi bilgisinin sonuca kısmi olarak aktarılmasında yardımcı olması ve model eğitiminde gradyanların yok olması problemini önleyerek öğrenmeyi verimli hale getirmesidir. Bu çalışmada PyTorch kütüphanesinin önceden eğitilmiş ResNet18<sup>2</sup> mimarisi transfer öğrenme olarak kullanılmış ve tam bağlantılı katmanında eğitimler gerçekleştirilmiştir.

### 2.3 Veri Kümesi

Deneyler için 10 sınıflı CIFAR10 [23] görüntü veri kümesi kullanılmıştır. CIFAR10 veri kümesi 60,000 adet 32x32 boyutunda görüntülerden oluşmakta ve bunun 50,000 adet kısmı model eğitimi için, 10,000 adet kısmı da doğrulama testleri için kullanılmaktadır. Veri kümesindeki bazı örnek görüntüler Şekil-1'de gösterilmiştir.

<sup>1</sup> <https://github.com/mv-lab/swin2sr>

<sup>2</sup> <https://pytorch.org>

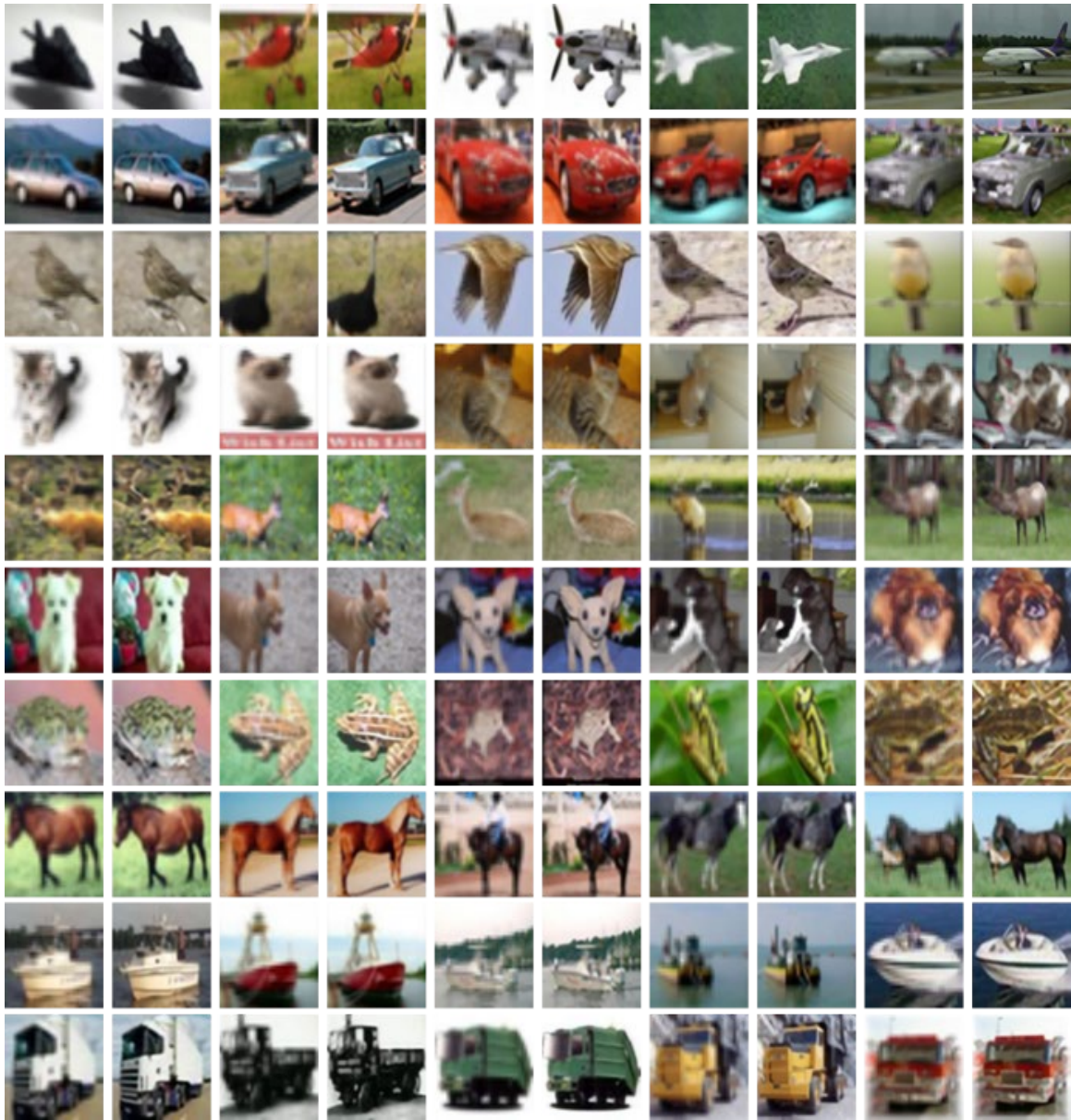
## 2.4 Çalışma Konfigürasyonu

Deneyler PyTorch kütüphanesi kullanılarak NVIDIA GeForce RTX 3050 Ti GPU destekli Intel Core i7-12700H işlemcili bir dizüstü bilgisayarda gerçekleştirilmiştir. Optimizasyon için 0.001 öğrenme parametresi ile Adam optimizasyon algoritması kullanılmıştır. Deneyler normal CIFAR10 veri kümesi ve Swin2SR ile dönüştürülmüş veri kümeleri üzerinde 50 epok için çalıştırılmıştır.

## 3. BULGULAR VE DEĞERLENDİRME (RESULTS AND EVALUATION)

### 3.1 Oluşturulan Görüntüler

Swin2SR algoritması kullanılarak çözünürlüğü ve keskinliği artırılmış bazı görüntüler Şekil-1'de gösterilmiştir.



**Şekil-1.** Görüntü çiftlerinden soldakiler, orijinal CIFAR-10 veri kümesindeki bazı görüntüler olup görüntü çiftlerinin sağ kısmındakiler ise soldaki görüntüye karşılık gelen Swin2SR Dönüştürme ile çözünürlüğü artırılmış ve keskinleştirilmiş görüntülerdir.

### 3.2 Değerlendirme Metrikleri

Normal CIFAR10 görüntüler ve Swin2SR dönüştürücü ile elde edilmiş görüntülerin ResNet18 modeli üzerindeki performanslarını değerlendirmek için doğruluk ve F1-Puanı metrikleri kullanılmıştır. İlgili metrikler aşağıda detaylandırılmıştır.

**Doğruluk:** Modelin vermiş olduğu kararların doğruluk oranını belirler ve Denklem-1'deki gibi hesaplanır:

$$\text{Doğruluk} = \frac{DP + DN}{DP + DN + YP + YN} \quad (1)$$

Denklem (1)'deki kısaltmalar; DP: doğru pozitif, DN: doğru negatif, YP: yanlış pozitif, YN: yanlış negatif olarak verilmiş kararların sayısını belirtmektedir. CIFAR10 ve Swin2SR dönüştürülmüş görüntülerin ResNet18 mimarisi üzerinde 50 epokluk eğitimi sırasında elde edilen eğitim doğrulukları Şekil-2'de gösterilmiştir.

**F1-Puanı:** Makine öğrenmesinde kullanılan bir performans ölçüm metriğidir. Hesaplamanın yapılabilmesi için iki bileşene ihtiyaç duyar: kesinlik (precision) ve duyarlılık (recall). Bu bileşenler kullanılarak dengeli bir performans ölçümü sağlanır.

*Kesinlik:* Modelin pozitif olarak belirlemiş olduğu kararların (ilgili etiketin atanması) doğruluğunu hesaplar.

$$\text{Kesinlik} = \frac{DP}{DP + YP} \quad (2)$$

*Duyarlılık:* Modelin pozitif örnekleri doğru olarak tanıyabilme oranını vermektedir.

$$\text{Duyarlılık} = \frac{DP}{DP + YN} \quad (3)$$

F1-Puanı kesinlik (Denklem-2) ve duyarlılık (Denklem-3) bileşenleri kullanılarak Denklem-4'teki gibi hesaplanır.

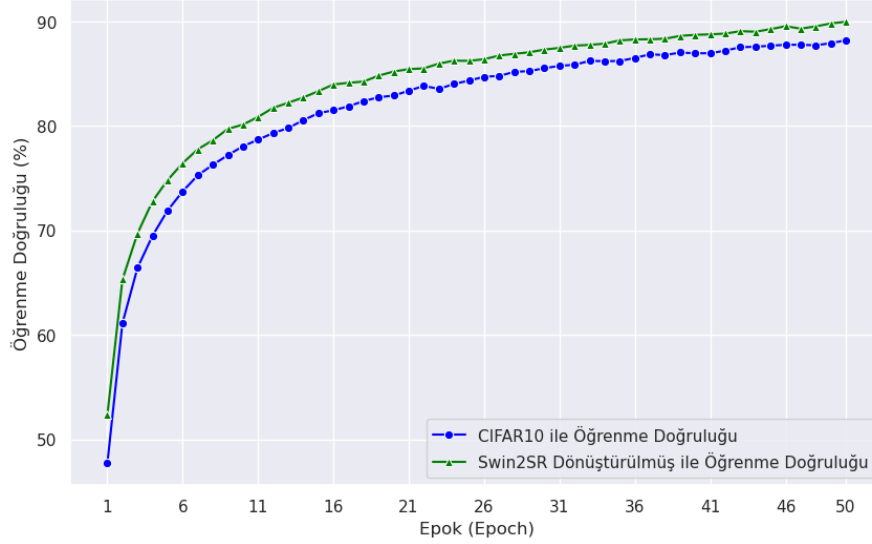
$$F_1 = \frac{2 * \text{Kesinlik} * \text{Duyarlılık}}{\text{Kesinlik} + \text{Duyarlılık}} \quad (4)$$

CIFAR-10 ve Swin2SR dönüştürülmüş veri kümelerinin ResNET18 modeli üzerinde 50 epok (epoch) eğitildikten sonra doğrulama (validation) kümesi üzerinde elde edilen doğruluk ve F1-Puanları Tablo-1'de gösterilmiştir.

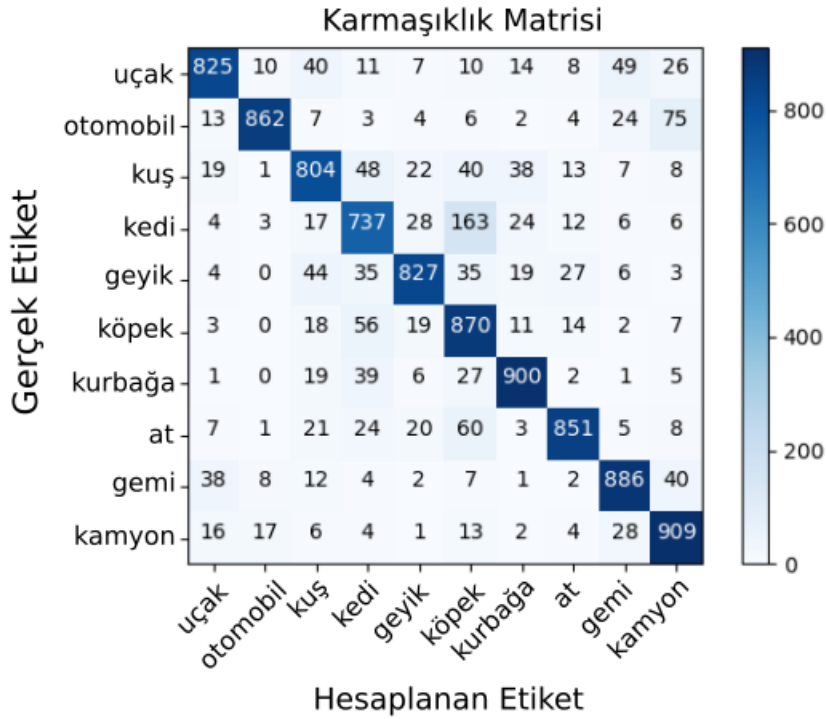
**Tablo 1.** Transfer öğrenme olarak ResNet18 mimari modeli kullanılarak doğrulama kümesi üzerinden elde edilmiş sonuçlar.

|   | <b>Doğruluk</b> | <b>F1-Puanı</b> |
|---|-----------------|-----------------|
| CIFAR10 Veri Kümesi                           | 0.85            | 0.85            |
| Swin2SR ile Dönüştürülmüş CIFAR10 Veri Kümesi | 0.87            | 0.87            |

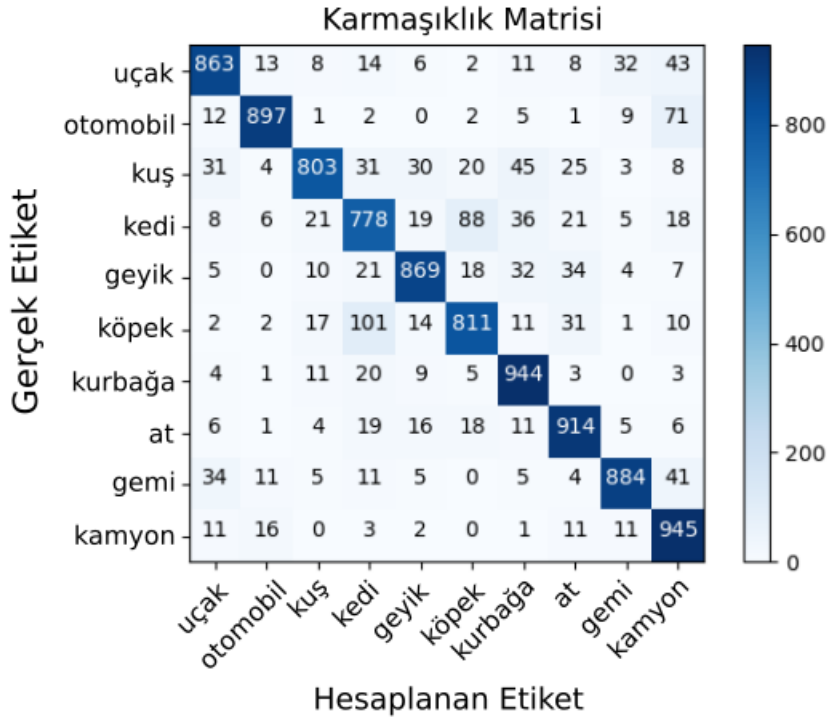
CIFAR10 ve Swin2SR ile dönüştürülmüş veri kümelerinin ResNet18 modeli ile sınıflandırma yapılırken sınıf bazındaki sınıflandırma sonuçları Şekil-3 ve Şekil-4'teki karmaşıklık matrislerinde gösterilmiştir. Her iki durumda kamyon sınıfı en yüksek oranda tanınırken, en düşük tanınma kedi sınıfında olmuştur.



Şekil-2. CIFAR10 veri kümesi ve Swin2SR ile elde edilmiş veri kümesinin ResNet18 modeli üzerinde 50 epokluk öğrenme doğruluğu grafiği.



Şekil-3. CIFAR-10 veri kümesinin doğrulama kısmı üzerinde ResNet18 modeli üzerindeki çalışması sonucu oluşan karmaşıklık matrisi.



**Şekil-4.** CIFAR10 veri kümesinin doğrulama kısmının Swin2SR dönüştürülmesiyle elde edilen veri kümesinin ResNet18 modeli üzerindeki çalışması sonucu oluşan karmaşıklık matrisi.

#### 4. SONUÇLAR (CONCLUSIONS)

Bu çalışmada görüntü çözünürlüğünün ve keskinliğinin artırılması ile görüntü sınıflandırma performansının değişimi incelenmiştir. Artık modül tabanlı ResNet18 mimarisi kullanılarak CIFAR10 görüntü veri kümesi üzerinde yapılan deneyler göstermiştir ki Swin tabanlı Swin2SR modeli ile görüntü kalitesinin iyileştirilmesi ile birlikte doğruluk ve F1-Puanı metriklerinde artış sağlanmaktadır. Bu sonuçlar bize görüntü sınıflandırma öncesi görüntü kalitesinin iyileştirilmesinin bir ön işlem olarak yapılmasının görüntü sınıflandırma performansını artıracaklarını göstermektedir. Bu yaklaşımın diğer problem alanlarında da kullanılabileceği düşünülmektedir. Özellikle uzaktan algılama alanında yapılan çalışmalarda uydu görüntülerinin çözünürlükleri ve keskinlikleri düşük olmaktadır. Bu düşük çözünürlüklü ve keskinliği az uydu görüntülerinde Swin2SR algoritması kullanılarak görüntü kalitesinin iyileştirilmesi ile birlikte uzaktan algılama çalışmalarındaki sınıflandırma performansının artırılacağı düşünülmektedir. Bu kapsamda farklı spektral uydu görüntüleri üzerinde Swin2SR ile görüntü kalitesinin artırılıp uydu görüntüleri üzerinde sınıflandırma çalışmalarının yapılması planlanmaktadır.

#### ÇIKAR ÇATIŞMASI (CONFLICT OF INTEREST)

Makale yazarları aralarında herhangi bir çıkar çatışması olmadığını beyan ederler.

#### KAYNAKLAR (REFERENCES)

- [1] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet Classification with Deep Convolutional Neural Networks, Advances in Neural Information Processing Systems, 25, 2012.
- [2] G. Huang, Z. Liu, L. Van Der Maaten, K. Q. Weinberger, Densely Connected Convolutional Networks, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4700-4708), 2017.

- [3] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 770-778), 2016.
- [4] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going Deeper with Convolutions, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1-9), 2015.
- [5] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, Mobilenets: Efficient Convolutional Neural Networks for Mobile Vision Applications, arXiv preprint arXiv:1704.04861, 2017.
- [6] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-scale Image Recognition, arXiv preprint arXiv:1409.1556, 2014.
- [7] S. J. Pan, Q. Yang, A Survey on Transfer Learning, IEEE Transactions on Knowledge and Data Engineering, 22(10), 1345-1359, 2009.
- [8] C. Shorten, T. M. Khoshgoftaar, A Survey on Image Data Augmentation for Deep Learning, Journal of Big Data, 6(1), 1-48, 2019.
- [9] H. Chen, Y. Pei, H. Zhao, Y. Huang, Super-resolution Guided Knowledge Distillation for Low-resolution Image Classification, Pattern Recognition Letters, 155, 62-68, 2022.
- [10] S. Hao, W. Wang, Y. Ye, E. Li, L. Bruzzone, A Deep Network Architecture for Super-resolution-aided Hyperspectral Image classification with Classwise Loss. IEEE Transactions on Geoscience and Remote Sensing, 56(8), 4650-4663, 2018.
- [11] F. M. Senalp, M. Ceylan, Effects of the Deep Learning-based Super-resolution Method on Thermal Image Classification Applications, Multimedia Tools and Applications, 81(7), 9313-9330, 2022.
- [12] J. D. Van Ouwerkerk, Image Super-resolution Survey, Image and Vision Computing, 24(10), 1039-1052, 2006.
- [13] K. Nasrollahi, T. B. Moeslund, Super-resolution: A Comprehensive Survey, Machine Vision and Applications, 25, 1423-1468, 2014.
- [14] Z. Wang, J. Chen, S. C. Hoi, Deep Learning for Image Super-resolution: A Survey, IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(10), 3365-3387, 2020.
- [15] C. Dong, C. C. Loy, K. He, X. Tang, Image Super-resolution Using Deep Convolutional Networks, IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(2), 295-307, 2015.
- [16] C. Tian, X. Zhang, J. C. W. Lin, W. Zuo, Y. Zhang, C. W. Lin, Generative Adversarial Networks for Image Super-resolution: A Survey, arXiv preprint arXiv:2204.13620, 2022.
- [17] Z. Lu, J. Li, H. Liu, C. Huang, L. Zhang, T. Zeng, Transformer for Single Image Super-resolution, In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 457-466), 2022.
- [18] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention is All You Need, Advances in Neural Information Processing Systems, 30, 2017.
- [19] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby, An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, arXiv preprint arXiv:2010.11929, 2020.

- [20] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, M. Shah, Transformers in Vision: A Survey, *ACM Computing Surveys (CSUR)*, 54(10s), 1-41, 2022.
- [21] M. V. Conde, U. J. Choi, M. Burchi, R. Timofte, Swin2SR: Swinv2 Transformer for Compressed Image Super-resolution and Restoration, In *European Conference on Computer Vision* (pp. 669-687), Cham: Springer Nature Switzerland, 2022.
- [22] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo, Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows, In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 10012-10022), 2021.
- [23] A. Krizhevsky, G. Hinton, Learning Multiple Layers of Features from Tiny Images, 2009.