

Improved Remote Sensing Image Classification using Axial Attention

Fatih Başatemur^{2,3} and Gökhan Bilgin^{1,3}

¹gbilgin@yildiz.edu.tr

²fatih.basatemur@std.yildiz.edu.tr

³Faculty Of Electrical & Electronics, Computer Engineering, Yildiz Technical University

Özetçe —ViT (Vision-Transformer), Convolutional Neural ağlara kıyasla umut verici performansı nedeniyle görüntü sınıflandırma görevlerinde popüler hale gelmiştir. Bu nedenle, birçok araştırmacı ViT modellerini hiper spektral görüntü (HSI) sınıflandırma görevlerine dahil etmeye çalışmış ancak tatmin edici bir performans elde edememiştir. Son yapılan çalışmalarda LIDAR ve HSI görüntü verilerini birleştirerek Multimodal Fusion Transformer (MFT) tabanlı yaklaşımlar popüler hale gelmiştir. Bu makalede, bu yaklaşımlardan biri olan Multimodal Fusion Transformer for Remote Sensing Image Classification çalışmasının geleneksel ViT yaklaşımlarında önerilen Multi-Head Cross Attention yerine daha başarılı ve güncel çalışmalardan Axial Attention modülünün farklı formları ve daha derin ardışıl encoder mimarileri için Residual Attention modülleri tanıtılmaktadır. Referans makaledeki çalışmanın encoder bloklarındaki gelişmeler incelendiğinden, çalışmanın diğer aşamalarına bağlı kalmamıştır.

I. GİRİŞ

İklim değişikliği ve çölleşme gibi durumlar, Uzaktan Algılama (UA) ve Yer Gözlemi'nin popüleritesinde büyük bir artışa neden olmuştur. Bu alanlarda yapılan çalışmalara örnek olarak arazi yüzeyi sınıflandırması, orman yönetimi, maden arama ve haritalama, nesne tespiti, çevresel izleme, kentsel planlama, biyoçeşitlilik koruma, felaket yanıtı ve yönetimi gibi görevler verilebilir. Bu görevlerin çoğu, hiper spektral görüntüleme cihazları gibi tek bir sensörden elde edilen veriler kullanılarak gerçekleştirilmiştir. Ayrıca, belirli nesneleri daha iyi tanımlamak ve ayırt etmek için LIDAR sensörlerinden yararlanılarak yükseklik bilgisine dayalı yüzey analizi yapılabilmektedir.

LIDAR, aynı spektral özelliklere ve farklı yüksekliklere sahip arazi örtüsündeki nesneleri ayırt etmemizi sağlayan yükseklik bilgisi sunar. Örneğin, HSI verileri su ve çimen gibi malzemeleri doğru bir şekilde ayırabilirken, aynı malzemelerden yapılan yolları ve çatıları ayırt etmek zordur. Bu nedenle, LIDAR'dan elde edilen yükseklik bilgisi, sınıflandırma amaçları için oldukça faydalıdır.

Son zamanlarda, Uzaktan Algılama verilerinin sınıflandırılmasında deep learning (DL) yöntemleriyle multimodal veri entegrasyonu büyük bir ilgi odağı haline gelmiştir. DL yöntemleri, büyük miktarda veriden spektral ve mekansal özellikleri etkili bir şekilde öğrenebilir. Ancak önceki çalışmalarda, bu tür CNN tabanlı modellerle yapılan füzyon analizi, orta ve uzun vadeli bağımlılıkları neredeyse hiç yakalayamaz. Bu durum, özellikle benzer spektral imzalara sahip birçok sınıfta bulunduğu verilerde performansın düşmesine ve spektral özelliklerin çıkarılmasında sorunlara neden olur.

Öte yandan, Recurrent Neural ağlar, HSI verilerinden spektral imzaları bant bant biriktirerek doğru bir şekilde modelleyebilir. Bu nedenle, spektral bantların sırası, uzun vadeli bağımlılıkların öğrenilmesi ve gradyan kaybı problemlerinin önlenmesi açısından önemlidir. Bununla birlikte, HSI verilerinin birçok örneğe sahip olması ve RNN'lerin modeli eşzamanlı olarak eğitememeleri, sınıflandırma performansını sınırlayan bazı zorluklar doğurur.

Referans alınan çalışmada, yukarıda bahsedilen sınırlamaları ele almak ve HSI, SAR, DSM, LiDAR gibi diğer multimodal veri kaynaklarını değerlendirmek için, transformerlar kullanan Multimodal Fusion Transformer (MFT) yaklaşımı Antonio Plaza ve arkadaşları tarafından önerilmiştir[1]. Yeni geliştirilen MFT, transformerların ve füzyon özelliklerini Remote Sensing sınıflandırma görevleri için kullanır. Ancak Cross Attention tabanlı temel ViT encoder bloklarının kullanılması halen attention çıkarımı için mevcut blokların gelişime açık olduğunu göstermektedir.

Bu çalışmada, önceden önerilen Multi Head Cross Attention bloklarının yerine kullanılabilecek Axial attention modüllerinin farklı formları olan self, cross ve transpose axial attention yapıları ve ardışıl encoder mimarileri için Residual Attention modülleri önerilmektedir.

Makalenin geri kalanında, referans alınan çalışmada kullanılan encoder bileşenlerinin geliştirilmesi için önerilen yaklaşımlar Bölüm II'de açıklanmaktadır. Daha sonra, eğitim ve test için kullanılan datasetler Bölüm III'de açıklanmakta ve Bölüm IV'de datasetler üzerinde yapılan eğitim denemelerinden elde edilen sonuçlar tartışılmaktadır. Son olarak makale, Bölüm V de içerik ve genel değerlendirmeleri yaparak sona erer.

II. ÇALIŞMALAR

İlk olarak referans yaklaşımın encoder yapısı A'da özetlenmektedir. Sonrasında B'de Axial Attention yaklaşımından, Bölüm C ve D'de de Axial Attention'un türevlerinden ve son bölümde ise derin encoder blokları için kullanılabilecek Residual bloklar ifade edilmektedir.

A. Multi-Head Cross Attention

Multi-head cross attention mekanizması, görüntü işleme modellerinin farklı bölgeler arasındaki ilişkileri anlamasına yardımcı olmak için kullanılan bir attention mekanizmasıdır [3]. Bu mekanizmaya, birden fazla head uygulanarak girdiye farklı dikkat dağılımları oluşturur.

Referans çalışmada, LiDAR ve HSI verileri arasındaki ilişkiyi modellemek için bu mekanizma kullanılmaktadır [1]. Ancak, çalışma bu yönden yeterli olsa da, mevcut inceleme alanı içerisinde hala geliştirilebilir ve sınırlı kalmaktadır.

Cross attention, HSI ve LiDAR verileri arasındaki ilişkiyi modellemek için eşleştirilmiş verilere ihtiyaç duyar. Ancak, bu tür veri setlerinin toplanması ve etiketlenmesi zor ve zaman alıcı olabilir. Bu nedenle, sınırlı veri durumunda Cross Attention'ın performansı düşebilir. Ayrıca, hem HSI hem de LiDAR verilerinin elde edilmesi genellikle maliyetlidir. Bu veri türlerini bir araya getirerek Cross Attention kullanmak, ekstra maliyetlere yol açabilir. Özellikle büyük veri setleri üzerinde çalışırken, işlem gücü ve depolama gereksinimleri artabilir.

Çalışmamızda, HSI ve LiDAR verilerini daha verimli bir şekilde analiz edebilecek diğer dikkat mekanizmalarına ve farklı formlarına odaklanıyoruz. Bu şekilde, daha iyi sonuçlar elde etmek için mevcut inceleme alanının sınırlılıklarını aşmayı hedefliyoruz.

B. Axial Attention

Axial attention, büyük boyutlu verilerin verimli bir şekilde işlenmesi için önerilen self-attention tabanlı bir dikkat mekanizmasıdır. Bu mekanizma, giriş verilerini birbirleriyle ilişkilendirmek ve önemli bilgileri vurgulamak amacıyla kullanılır. Axial attention farklı boyut/kanalların (örneğin, satır ve sütun) özellikleri arasındaki ilişkileri yakalamak için özel olarak tasarlanmıştır[2].

Axial attention'ın çalışma adımları aşağıdaki gibi ifade edilebilir:

- Verinin ayrıştırılması: İlk adımda, girdi veri seti önceden belirlenmiş boyutlara göre ayrıştırılır. Örneğin, bir görüntüde her pikselin satır ve sütun bilgileri ayrı ayrı ele alınır. Bu ayrıştırma işlemi, veriyi boyutsal olarak parçalamak için uygun bir işlemi ifade eder.
- Boyutsal dikkat oluşturma: Ayrıştırılan veri seti üzerinde, her boyut (örneğin, satır ve sütun) için ayrı attention mekanizmaları oluşturulur. Bu mekanizmalar, boyuta özgü ilişkileri modellemek ve dikkat dağılımlarını belirlemek için kullanılır.
- Attention matrislerinin hesaplanması: Satır ve sütun dikkat mekanizmaları kullanılarak, her boyutta ilişki matrisleri hesaplanır. Bu matrisler, boyutlar arasındaki ilişkileri temsil eder ve her bir hücredeki değer, ilgili boyutlar arasındaki ilişkinin gücünü ifade eder.
- Attention verilerini birleştirme: Ağırlıklandırılmış attention matrisleri kullanılarak, orijinal girdi verileri boyutsal olarak birleştirilir. Bu işlem, ilişkili verilerin bir araya getirilmesini sağlar ve daha kapsamlı bir temsil elde edilir.
- Sonuç üretimi: Birleştirilmiş attention verileri, Feed-Forward network'e girdi olarak verilir ve stage çıktısı elde edilmiş olunur.

Axial attention yapısının Cross Attention yerine tercih edilmesi ve HSI - LiDAR verileriyle yapılan sınıflandırma çalışmalarına olan getirileri aşağıdaki gibi listeleyebilir:

- İkili ilişkinin yakalanması: HSI ve LiDAR verileri arasındaki ilişkiler genellikle karmaşık olabilir. Cross attention, ikili ilişkileri vurgulamada daha sınırlı olabilirken, Axial attention farklı boyutlardaki (örneğin, satır ve sütun) ilişkileri daha etkili bir şekilde yakalayabilir. Bu, HSI ve LiDAR verileri arasındaki karmaşık ilişkileri daha iyi modellemek için Axial attention'ın daha uygun olabileceğini gösterir.
- Veri boyutlarının farklılığı: HSI ve LiDAR verileri genellikle farklı boyutlara sahiptir. HSI verileri genellikle spektral kanallardan oluşurken, LiDAR verileri 3D nokta bulutlarından oluşur. Cross attention, farklı boyutlardaki veriler arasındaki ilişkileri modellemekte daha zorlanabilir. Ancak Axial attention, farklı boyutlardaki ilişkileri özellikle vurgulamak için daha iyi bir seçenek olabilir.
- Veri setinin özellikleri: HSI ve LiDAR verileri genellikle farklı özellikleri temsil eder. HSI verileri spektral özellikleri içerirken, LiDAR verileri yükseklik ve geometrik özellikleri içerir. Axial attention, bu farklı özellikleri daha iyi bir şekilde ele alabilir ve sınıflandırma modeline daha ayrıntılı bilgiler sağlayabilir.
- Veri setinin büyüklüğü: Cross attention, daha fazla hesaplama gücüne ihtiyaç duyabilir ve daha büyük veri setlerinde maliyetli olabilir. Öte yandan, Axial attention, veriyi boyutsal olarak ayrıştırarak ve boyutsal dikkat mekanizmaları kullanarak daha verimli bir hesaplama sağlayabilir. Bu nedenle, daha büyük veri setleri üzerinde çalışırken Axial attention'ın avantajlı olabileceği düşünülebilir.

HSI ve LiDAR verilerini kullanarak sınıflandırma yapmak için encoder olarak Cross yerine Axial attention'ın tercih edilmesi, daha verimli hesaplama sağlayabilir ve daha iyi özellik temsili sunabilir. Bu çalışmada, MFT modeller için Axial attention yapısının sınıflandırma performansını artırması ve daha iyi sonuçların elde edilmesine katkısı olacağı düşünülmektedir.

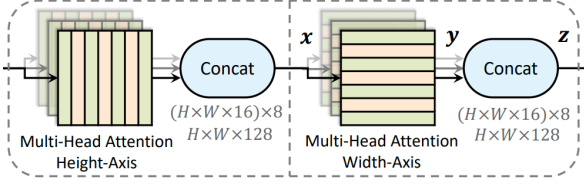
C. Cross Axial Attention

Bilindiği üzere, Axial attention mekanizması, kanal/boyutlar arasındaki ilişkileri modellerken self-attention yerine cross-attention mekanizmasını kullanır. LiDAR ve HSI verilerinin ikili ilişkilerini modellerken cross attention mekanizması, self attention'a kıyasla daha başarılı bir şekilde farklı öğeler arasındaki ilişkileri yakalayabilir. Bu nedenle, özellikle HSI ve LiDAR gibi farklı modaliteler arasındaki ilişkileri anlamak için cross attention mekanizması daha uygun olabilir. Ayrıca, cross attention, farklı özellikleri temsil eden verilerin entegrasyonunda daha etkili olabilir ve bu verilerin birbirleriyle etkileşimini dikkate alarak daha kapsamlı bir temsil sağlayabilir. Örneğin, spektral görüntü özellikleri ile LiDAR yükseklik bilgileri arasındaki ilişkileri daha iyi modelleyebilir.

Bu avantajlar, HSI ve LiDAR verileriyle yapılan sınıflandırma modellerinde cross attention mekanizmasının Axial attention'a tercih edilme nedenlerini ortaya koymaktadır. Bu nedenle, axial attention encoder bloklarında self-attention yerine cross-attention kullanılmasının başarıyı arttırabileceği düşünülmektedir.

D. Transpose Axial Attention

Huiyu Wang ve diğer araştırmacılar tarafından önerilen Stand-Alone Axial-Attention [4] yaklaşımı, girdiler arasındaki yapısal bütünlüğü modellemek amacıyla ortaya atılmıştır. Bu yaklaşımda, Transpose işlemi encoder blokları arasında kullanılarak ilişki modellemesinin başarımını artırmayı hedeflemektedir.



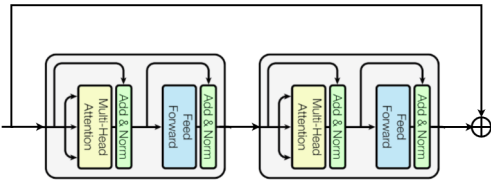
Şekil 1: Stand-Alone Axial-Attention yaklaşımında kullanılan transpose attention blokları ifade edilmektedir.

Encoder bloklarının implementasyon işlemi, attention girdi ve çıktılarına transpose işlemi uygulanarak gerçekleştirilir. Bu durumda, ilk bloğun axial attention çıktısı, bir sonraki bloğa transpose edilerek input olarak verilir. Sonraki bloğa çıktıyı input olarak verebilmek için çıktı tekrar transpose edilir.

Bu yaklaşım ile attention modülleri arasındaki boyutsal ilişkilerin özel olarak vurgulanması ve daha etkili yakalanması hedeflenmektedir. Bundan dolayı mevcut çalışmada MFT çalışmaları için uygulanabilir ve incelemeye değer görülmüştür.

E. Residual Skip Connection

Skip connection, derin sinir ağlarının performansını ve yakınsamasını iyileştirmek için yaygın olarak kullanılan bir tekniktir [7]. Bu teknik, lineerite kaybını hafifletmek için sinir ağı katmanları boyunca doğrusal bileşenleri iletmeyi amaçlar. Böylece gradient vanishing ve gradient exploding problemlerinin de önüne geçilmiş olunur.



Şekil 2: Residual değerler, iki sonraki bloğun çıktısı ile toplanmaktadır.

Bu çalışmada, encoder blokları arasında değerlerin ileri yönde, bir önceki bloğun girdisinin ve iki sonraki bloğun çıktısına eklenmesi ile uygulanmıştır. Bu yaklaşım kanalları ve head'ler arasındaki ilişkinin yakınsamasına yardımcı olabilir. Sonraki çalışmalarda blok sayılarının derinleştirilmesi ile yaklaşımın değerlendirilmesi düşünülmektedir.

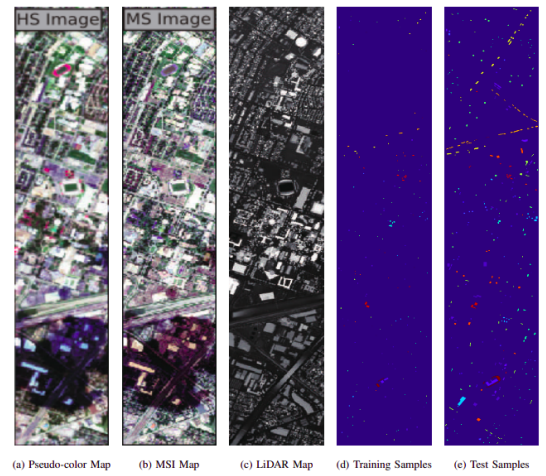
III. HSI VE LIDAR DATASET

Bu bölümde, önerilen attention mekanizmalarının performansını değerlendirmek için üç farklı HSI ve bunlara ilişkin çoklu modalite veri kaynaklarını (LiDAR, MS, SAR ve DSM)

ele alıyoruz. University of Houston (UH), Trento ve MUUFL veri setleri bu çalışmada kullanılan veri kümeleri arasında yer almaktadır.

A. University of Houston

University of Houston veri seti, Compact Airborne Spectrographic Imager (CASI) tarafından toplanmıştır. Dataseti HSI (Hiperspektral Görüntü), MS (Multi Spektral) görüntü ve bir LiDAR görüntüsünden oluşmaktadır. Tüm görüntüler 340x1905 pikselden oluşurken, HSI 144 bant ve MS görüntü 8 spektral bant içermektedir. Bu veri setinin piksel başına 2.5 metre mekansal çözünürlüğü ve 0.38 - 1.05 μm arasında dalga boyu bulunmaktadır. Zemin gerçekliği 15 farklı arazi örtüsü ve kullanım sınıfını içermektedir. Ayrıca, 15 arazi türü sınıfının örnekleri sabit boyutlu eğitim ve test örneklerine ayrılmıştır. Şekil 3, arazi örtüsü ve kullanımının 15 farklı kategorisini ve bunlarla ilişkili eğitim ve test örneklerini göstermektedir.



Color	Land cover	Train	Test	Color	Land cover	Train	Test
Blue	Background	662013	652648	Blue	Grass-healthy	198	1053
Green	Grass-stressed	190	1064	Green	Grass-synthetic	192	505
Yellow	Tree	188	1056	Yellow	Soil	186	1056
Red	Water	182	143	Red	Residential	196	1072
Purple	Commercial	191	1053	Purple	Road	193	1059
Orange	Highway	191	1036	Orange	Railway	181	1054
Light Blue	Parking-lot1	192	1041	Light Blue	Parking-lot2	184	285
Dark Blue	Tennis-court	181	247	Dark Blue	Running-track	187	473

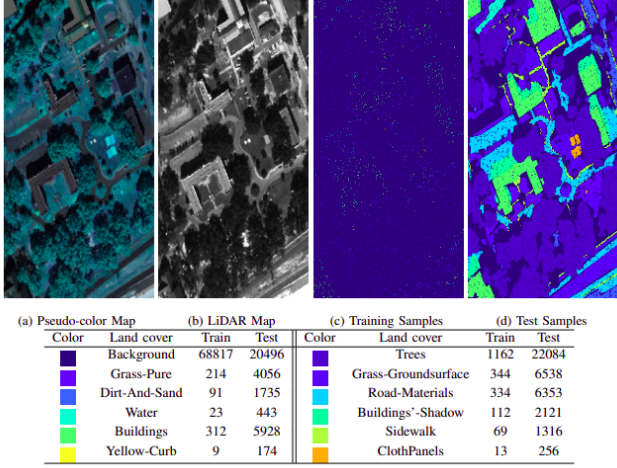
Şekil 3: (a) HSI verilerinden elde edilen Pseudo-color görüntü, sırasıyla bantlar 64, 43 ve 22 kullanılarak oluşturulmuştur. (b) MSI verilerinden elde edilen Grayscale görüntü, (c) LiDAR verilerinden elde edilen Grayscale görüntü, (d) Eğitim örneği değeri ve (e) Test örneği değeri. Tablo, University of Houston (UH) datasetinde bulunan yüzey türlerini ve birbirinden farklı eğitim ve test örneklerinin sayısını göstermektedir.

B. Muufl

MUUFL Gulfport dataseti, ROSIS sensörü kullanılarak Kasım 2010'da Mississippi Üniversitesi kampüsü üzerinde toplanmıştır [5], [6]. Bu veri setinin HSI'sında 325x220 piksel ve 72 spektral bant bulunmaktadır. Bu veri setinin LiDAR görüntüsü, 2 raster yükseklik verisini içermektedir. Gürültü nedeniyle 8 başlangıç ve son bant kaldırılmış ve toplamda 64 bant elde edilmiştir. Dataseti, 53687 piksel içeren 11 kentsel arazi örtüsü sınıfını tasvir etmektedir. Şekil 5, her sınıftan rastgele seçilen %5 örneğin dağılımını göstermektedir.

Dataset	TRENTO			MUUFL			HOUSTON		
	OA	F1	KAPPA	OA	F1	KAPPA	OA	F1	KAPPA
MFT	97.28185	0.93333	96.35799	90.85169	0.747272	87.94099	84.96351	0.851333	83.72075
Axial Attention	98.02687	0.953333	97.35718	91.02031	0.742727	88.07428	85.77519	0.85933	84.5894
Cross Axial Attention	97.74451	0.941666	96.97706	91.58693	0.765454	88.82153	84.26662	0.84533	82.97592
Transpose Axial Attention	97.48596	0.928333	96.63119	91.03795	0.74	88.11505	85.11109	0.835333	83.87818
Residual Skip Connection	96.95186	0.913333	95.91451	91.26146	0.749090	88.42401	85.63581	0.858	84.41465

Şekil 4: Trento, Muufl ve Houston datasetleri kullanılarak yapılan eğitim sonucunda elde edilen OA, F1 ve Kappa skorları gözlemlenmektedir.



Şekil 5: (a) HSI verilerinden elde edilen True-color görüntü, sırasıyla bantlar 40, 20 ve 10 kullanılarak oluşturulmuştur. (b) LiDAR verilerinden elde edilen Grayscale görüntü (c) MUUFL sahnesinin gerçek değerleridir. Tablo, MUUFL datasetindeki yüzey türlerini ve rastgele seçilen %5 eğitim ve kalan %95 test örneklerinin sayısını göstermektedir.

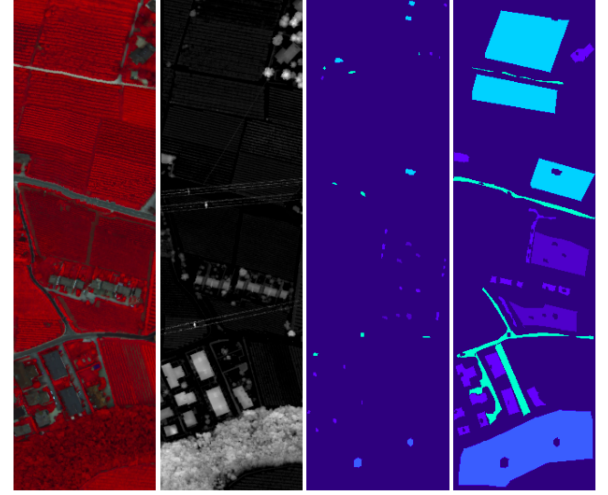
C. Trento

AISA Eagle sensörleri, İtalya'nın Trento şehrinin güneyindeki kırsal bölgelerde HSI verilerini toplamak için kullanılmıştır. Optech ALT M 3100EA sensörleri ise LiDAR verilerini toplamıştır. Her bir HSI'da 63 bant bulunmakta olup dalga boyu 0.42-0.99 μm arasında değişmektedir. LiDAR verilerinde ise yükseklik bilgisi sağlayan bir piksel bulunmaktadır. Spektral çözünürlük 9.2 nm, mekansal çözünürlük ise piksel başına 1 metredir. Sahne, birbirleriyle örtüşmeyen 6 bitki örtüsü sınıfını içermekte olup piksel sayısı 600×166 'dır. Şekil 6'da her sınıf için eğitim ve test örnekleri hakkında verilmektedir.

IV. DENEMELER

A. Dataset ve Değerlendirme Metrikleri

Çalışmada University of Houston, Muufl ve Trento datasetleri üzerinden denemeler yapılmıştır. Houston datasetinde toplamda 15 sınıf için HSI ve LiDAR verilerinden eğitim kümesine 2832 ve test kümesine 12192 örnek ayrılmıştır. Muufl datasetinde toplamda 11 sınıf için HSI ve LiDAR verilerinden eğitim kümesine 2683 ve test kümesine 51004 örnek ayrılmıştır. Trento datasetinde ise toplamda 6 sınıf için HSI ve LiDAR verilerinden eğitim kümesine 819 ve test kümesine 29395 örnek ayrılmıştır.



Şekil 6: (a) HSI verilerinden bantlar 40, 20 ve 10 kullanılarak oluşturulan True-color görüntü. (b) LiDAR verilerinden elde edilen Grayscale görüntü. (c) Eğitim örneği değerleri. (d) Test örneği değerleri. Tablo, Trento datasetindeki yüzey türlerini ve birbirinden farklı eğitim ve test örneklerinin sayısını göstermektedir.

Mevcut çalışmada UH, Muufl ve Trento datasetleri kullanılarak belirtilen spesifikler ile yapılan 40 epoch eğitim sonrasında test dataseti üzerinden elde edilen Overall Accuracy (OA) [8], F1 [9] ve Kappa [10] skorlarının sonuçları Şekil 4'de ifade edilmiştir.

V. SONUÇ VE TARTIŞMA

Şekil 4'teki tabloya bakıldığında, referans alınan MFT çalışmasına göre kullanılan veri kümeleri incelendiğinde, Axial Attention mekanizmasının Trento veri seti için daha yüksek OA, F1 ve Kappa skorlarına sahip olduğu görülmektedir. Muufl veri seti için ise Cross Axial Attention yaklaşımının daha iyi sonuçlar verdiği ve Houston veri seti için Axial ve Residual yaklaşımlarının daha başarılı olduğu söylenebilir.

Sonuç olarak, bu çalışma HSI ve LiDAR gibi büyük verilerin işlendiği çalışmalarda Axial Attention yaklaşımının Multimodal Fusion Transformer tabanlı yöntemlere kıyasla daha başarılı bir sınıflandırma sağladığını göstermektedir.

Sınıflar	MFT			Axial Attention			Cross Axial Attention			Transpose Axial Attention			Residual Skip Connection		
	precision	recall	F1	precision	recall	F1	precision	recall	F1	precision	recall	F1	precision	recall	F1
Apples	0.98	0.98	0.98	1.00	0.98	0.99	0.99	0.97	0.98	1.00	0.96	0.98	1.00	0.95	0.97
Buildings	0.82	1.00	0.90	0.86	1.00	0.93	0.86	1.00	0.92	0.85	1.00	0.92	0.82	1.00	0.90
Ground	0.83	0.83	0.83	0.82	0.95	0.88	0.79	0.89	0.84	0.70	0.86	0.77	0.67	0.82	0.74
Woods	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Vineyard	0.99	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	0.99	1.00	1.00	0.99	1.00	0.99
Roads	0.99	0.80	0.89	0.99	0.86	0.92	0.99	0.84	0.91	0.99	0.83	0.90	0.99	0.79	0.88

Şekil 7: Tabloda, Trento test datasetinde bulunan 6 sınıf için MFT ve önerilen yaklaşımların elde ettiği precision, recall ve F1 skorları gösterilmektedir.

Sınıflar	MFT			Axial Attention			Cross Axial Attention			Transpose Axial Attention			Residual Skip Connection		
	precision	recall	F1	precision	recall	F1	precision	recall	F1	precision	recall	F1	precision	recall	F1
Trees	0.97	0.95	0.96	0.96	0.97	0.97	0.96	0.98	0.97	0.96	0.97	0.97	0.97	0.97	0.97
Grass Pure	0.78	0.86	0.82	0.81	0.84	0.82	0.81	0.84	0.83	0.79	0.87	0.82	0.81	0.85	0.83
Grass Groundsurface	0.83	0.87	0.85	0.82	0.86	0.84	0.87	0.84	0.85	0.84	0.83	0.83	0.84	0.84	0.84
Dirt And Sand	0.85	0.95	0.90	0.84	0.96	0.90	0.91	0.90	0.91	0.86	0.95	0.90	0.82	0.97	0.89
Road Materials	0.88	0.95	0.92	0.88	0.95	0.92	0.89	0.96	0.92	0.89	0.95	0.92	0.89	0.95	0.92
Water	0.77	0.72	0.74	0.87	0.76	0.81	0.80	0.79	0.79	0.88	0.75	0.81	0.84	0.84	0.84
Buildings Shadow	0.91	0.78	0.84	0.93	0.78	0.85	0.90	0.84	0.87	0.87	0.86	0.86	0.87	0.86	0.86
Buildings	0.95	0.97	0.96	0.95	0.97	0.96	0.97	0.96	0.97	0.94	0.98	0.96	0.95	0.98	0.96
Sidewalk	0.60	0.35	0.45	0.80	0.19	0.30	0.66	0.44	0.52	0.74	0.17	0.28	0.79	0.20	0.33
Yellow Curb	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ClothPanels	0.89	0.70	0.78	0.94	0.70	0.80	0.90	0.70	0.79	0.92	0.69	0.79	0.85	0.75	0.80

Şekil 8: Tabloda, Muufl test datasetinde yer alan 11 sınıf için MFT ve önerilen yaklaşımların elde ettiği precision, recall ve F1 skorları bulunmaktadır.

Sınıflar	MFT			Axial Attention			Cross Axial Attention			Transpose Axial Attention			Residual Skip Connection		
	precision	recall	F1	precision	recall	F1	precision	recall	F1	precision	recall	F1	precision	recall	F1
Healthy grass	0.88	0.91	0.89	0.99	0.81	0.89	1.00	0.82	0.90	1.00	0.82	0.90	0.92	0.83	0.87
Stressed grass	1.00	0.84	0.91	1.00	0.83	0.91	1.00	0.83	0.91	0.80	0.97	0.87	0.92	0.91	0.91
Synthetic grass	0.45	0.99	0.62	0.51	0.96	0.67	0.36	0.97	0.52	0.64	0.98	0.78	0.86	0.94	0.90
Trees	0.96	0.98	0.97	1.00	0.89	0.94	0.95	0.97	0.96	0.95	0.98	0.96	0.88	0.96	0.92
Soil	1.00	1.00	1.00	0.99	1.00	0.99	1.00	0.99	1.00	1.00	1.00	1.00	0.99	0.98	0.99
Water	0.62	1.00	0.77	0.76	1.00	0.86	0.67	0.85	0.75	0.36	0.96	0.52	0.96	0.94	0.95
Residential	0.83	0.82	0.83	0.65	0.90	0.76	0.94	0.85	0.89	0.81	0.82	0.82	0.71	0.89	0.79
Commercial	0.82	0.73	0.77	0.88	0.75	0.81	0.87	0.67	0.76	0.90	0.69	0.78	0.95	0.73	0.82
Road	0.90	0.76	0.83	0.88	0.85	0.86	0.86	0.83	0.85	0.85	0.77	0.81	0.79	0.85	0.82
Highway	0.85	0.58	0.69	0.95	0.59	0.72	0.86	0.55	0.67	0.81	0.59	0.69	0.94	0.43	0.59
Railway	0.90	0.84	0.87	0.87	0.93	0.90	0.80	0.87	0.83	0.99	0.80	0.88	0.78	0.93	0.85
Parking Lot 1	0.78	0.87	0.82	0.86	0.89	0.87	0.79	0.89	0.84	0.80	0.95	0.87	0.83	0.92	0.88
Parking Lot 2	0.87	0.81	0.84	0.76	0.82	0.79	0.92	0.78	0.84	0.72	0.84	0.77	0.65	0.92	0.76
Tennis Court	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.98	0.99	0.99	1.00	0.99	0.93	0.82	0.87
Running Track	0.95	0.98	0.96	0.89	0.96	0.92	0.95	1.00	0.97	0.92	0.87	0.89	0.90	1.00	0.95

Şekil 9: Tabloda, Houston test datasetinde yer alan 15 sınıf için MFT ve önerilen yaklaşımların precision, recall ve F1 skorları sunulmaktadır.

KAYNAKLAR

- [1] Roy, S.K., Deria, A., Hong, D., Rasti, B., Plaza, A., Chanussot, J. (2022). Multimodal Fusion Transformer for Remote Sensing Image Classification. arXiv preprint arXiv:2203.16952. Retrieved June 18, 2023
- [2] Ho, J., Kalchbrenner, N., Weissenborn, D., Salimans, T. (2019). Axial Attention in Multidimensional Transformers. CoRR, abs/1912.12180.
- [3] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is all you need. arXiv preprint arXiv:1706.03762.
- [4] Wang, H., Zhu, Y., Green, B., Adam, H., Yuille, A. L., Chen, L.-C. (2020). Axial-DeepLab: Stand-Alone Axial-Attention for Panoptic Segmentation. CoRR, abs/2003.07853.
- [5] P. Gader, A. Zare, R. Close, J. Aitken, and G. Tuell, "Muuffl gulfport hyperspectral and lidar airborne data set," Univ. Florida, Gainesville, FL, USA, Tech. Rep. REP-2013-570, 2013.
- [6] X. Du and A. Zare, "Scene label ground truth map for muuffl gulfport data set," Dept. Elect. Comput. Eng., Univ. Florida, Gainesville, FL, USA, Tech. Rep. 2017.
- [7] He, K., Zhang, X., Ren, S., Sun, J. (2015). Deep Residual Learning for Image Recognition. CoRR, abs/1512.03385.
- [8] Sokolova, M., Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. Information Processing Management, 45(4), 427-437.
- [9] Sokolova, M., Japkowicz, N., Szpakowicz, S. (2006). Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation. AI 2006: Advances in Artificial Intelligence, 101-113.
- [10] Carletta, J. (1996). Assessing agreement on classification tasks: the kappa statistic. Computational Linguistics, 22(2), 249-254.