



Generating Ground Truth Images Using GAN (Generative Adversarial Network)

Elif Işılav Ünlü and Ahmet Çınar

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

September 28, 2022



ICIAS

<https://www.icias.net/>

GAN (Generative Adversarial Network) Kullanarak Temel Gerçeklik Görüntüleri Üretme

Elif Işılay ÜNLÜ^{1*}, Ahmet ÇINAR²

¹Bilgisayar Mühendisliği, Fırat Üniversitesi, TÜRKİYE
ORCID ID 0000-0002-0236-5527

²Bilgisayar Mühendisliği, Fırat Üniversitesi, TÜRKİYE
ORCID ID 0000-0001-5528-2226

*(eiunlu@firat.edu.tr)

Özet – Görüntü segmentasyonu görüntü üzerindeki her bir pikselin sınıflandırılmasıdır ve görüntü üzerinde bulunan tüm nesne alanlarının bir haritasını oluşturmaktır. Segmentasyon, biyomedikal, çevre, coğrafya vb. birçok alanda görüntülerin analizinde yaygın olarak kullanılmaktadır. Segmentasyon için kaynak görüntülerin yanı sıra bu görüntülere karşılık gelen temel gerçeklik görüntüleri mevcut olmalıdır. Ancak temel gerçeklik bilgilerini manuel olarak elde etmek çok fazla çaba ve zaman gerektirir. Bu çalışmada amacımız derin öğrenme tabanlı, uydu görüntülerinden temel gerçeklik görüntülerini üretebilen bir sistem geliştirmektir. Uydu görüntüleri GAN(Generative Adversarial Network) mimarisi kullanılarak eğitilmiştir. Model girdi olarak uydu görüntüsünü alır ve çıktı olarak hedef görüntüyü (google harita görüntüsü) oluşturur. Önerilen sistem, belirli bir uydu görüntüsüne karşılık gelen google harita görüntüsünü ve beklenen google harita görüntüsünün çıktısını verir. Benzerlik görsel olarak analiz edilebilmektedir.

Anahtar Kelimeler – segmentasyon, uydu görüntüsü, Görüntüden görüntüye dönüşüm, GAN, Pix2Pix

I. GİRİŞ

Görüntü segmentasyonu, görüntü üzerinde aynı özellikteki pikselleri parçalar halinde bir araya toplamaktır, görüntüyü belli bölgelere ayırmaktır. Görüntü içerisinde benzer parlaklıklar olabilir ve bu parlaklıklar ilgili görüntünün farklı bölgelerindeki nesnelere temsil edebilir. Görüntü içerisinde aynı parlaklıklara sahip nesne parçacıklarının belirlenmesi, sınıflandırma ve tanılama amacı için segmentasyon kullanılabilir [1]. Segmentasyon, biyomedikal, çevre, coğrafya vb. birçok alanda görüntülerin analizinde yaygın olarak kullanılmaktadır.

Görüntü segmentasyonu işleminde modelin eğitimi ve performans ölçümü için kaynak görüntülerine karşılık gelen hedef görüntüler bulunmalıdır. Bu hedef görüntüler temel gerçeklik(ground truth) olarak adlandırılmaktadır. Temel gerçeklik görüntüleri modelin tahmin etmesi istenilen gerçek verilerdir. Temel gerçeklik görüntülerini, alanında uzman bir veya daha fazla

kişi tarafından gerçekleştirilen ‘gerçek ve doğru’ bölümler şeklinde ifade edebiliriz.

Temel gerçeklik görüntülerini manuel olarak elde eden birçok yazılım aracı bulunmaktadır. Ancak temel gerçeklik bilgilerini manuel olarak elde etmek çok fazla çaba ve zaman gerektirir. Bu yazıda amacımız uydu görüntülerinden temel gerçeklik görüntülerini üretebilen derin öğrenme tabanlı, otomatik bir sistem geliştirmektir. Uydu görüntüsünü ilgili haritaya dönüştürmek için görüntüden görüntüye çeviri kullanılır. Girdi olarak belirli bir yakınlaştırma düzeyinde bir uydu görüntüsünü alan, o bölge için ilgili insan tarafından okunabilir haritaları oluşturmak için Üretken çekişmeli ağ, Koşullu rakip ağlar ve Ortak Varyasyonlu Otomatik kodlayıcılar gibi görüntüden görüntüye çeviriler için farklı teknikler kullanılır. Önerdiğimiz sistem Generative Adversarial Network(GAN) mimarisi ile uydu görüntülerine karşılık gelen temel gerçeklik görüntülerini üretmektedir.

GAN olarak adlandırılan bu model, yüksek çözünürlüklü görüntü elde etmek, veri setini genişletmek, resim boyamak, görüntüden görüntüye çeviri vb. işlemler için kullanılan popüler bir yaklaşımdır. GAN klasik derin ağ mimarilerinden farklı olarak birbirine zıt şekilde çalışan iki nöral ağdan oluşmaktadır. Bu ağlardan biri üretici(generator) diğeri ayırıcı(discriminator) ağdır. GAN bu iki ağın çekişmeli olarak çalışması ile öğrenme işlemini gerçekleştirir[2]. cGAN, DCGAN, InfoGAN, SGAN, ACGAN, WGAN-GP ve LSGAN gibi çeşitli GAN modeli bulunmaktadır.

A. İlgili Çalışmalar

Görüntü sentezi ve görüntü çevirisi, bilgisayarla görü alanında her zaman önemli bir araştırma alanı olmuştur. Üretken çekişmeli ağlar kullanılarak görüntüden görüntüye çeviri alanında çok sayıda önemli çalışma yapılmıştır. Son zamanlarda, GAN'lar görüntü oluşturmada [3],[4], görüntü çevirisinde [5],[6], süper çözünürlüklü görüntüleme [7] ve yüz görüntü sentezinde [8] dikkate değer sonuçlar göstermiştir. GAN'ın [2] üstün performansının arkasındaki ilk neden, üreticinin temel gerçek görüntülerinden ayırt edilemeyen sahte örnekler oluşturmasını sağlayan çekişmeli kayıptır. Koşullu Çekişmeli Ağlar [9], üretici ve ayırıcıyı sınıf etiketleri gibi bazı ekstra bilgiler üzerinde koşullandırarak GAN yapısını genişletir. Son yıllarda, çeşitli yöntemlerle görüntüden görüntüye çeviri alanında dikkate değer sonuçlar elde edilmiştir. Mevcut yöntemler denetimsiz ve denetimli olmak üzere iki kategoride sınıflandırılabilir. Pix2Pix [5] gibi denetimli yöntemler, Koşullu Çekişmeli Ağları kullanarak ve çekişme kaybını L1 kaybıyla birleştirir. Koşullu GAN'lar Deng ve arkadaşları tarafından [10] havai uyu görüntülerinden konumların yer seviyesindeki görünümünü oluşturmak için kullanılmıştır. Costea ve ark. [11], yolları ve kavşakları tespit etmek için çift atlamalı GAN'ı (DH-GAN) önerdi ve bunu, düzleştirme tabanlı bir grafik optimizasyon prosedürü uygulayarak en iyi kapsayan yol grafiğini bulmak için kullandı. Yakın zamanda Zhu ve arkadaşları tarafından önerilen CycleGAN mimarisi, uyu görüntülerinde harita çevirisi için değerlendirilmiştir[12]. Xu ve arkadaşları ayrıca, GAN'ları konumun uyu görüntüsüne bağlı olarak, sahte uyu görüntüleri ve temel gerçeğin sahte görüntülerini oluşturmak için kullanmıştır[13].

II. MATERYAL VE YÖNTEM

A. Veri Seti

Önerilen yöntemde, New York'un uyu görüntüleri ve bunlara karşılık gelen temel gerçeklik görüntüleri olmadığı için Google harita görüntülerinden oluşan bir veri seti kullanılmıştır[5]. Veri setinin 1096'sı eğitim, 1098'i test olarak gruplandırılmıştır. Modeli eğitmek için kullanılan örnek bir uyu görüntüsü ve google harita görüntüsü Şekil 1'de verilmiştir. Her bir görüntü 1200x600 boyutlarında jpeg formatındadır. Eğitim veri setindeki görüntüler yeniden ölçeklendirilerek 256x256 boyutlarına dönüştürülmüştür.



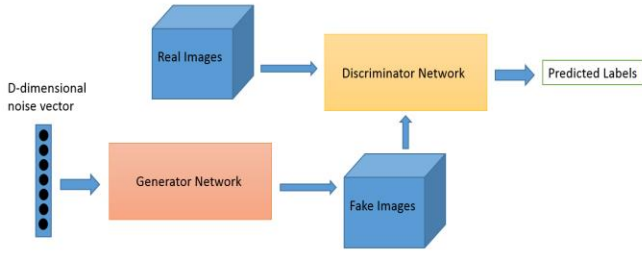
Şekil 1. Uyu görüntüsü ve karşılık gelen google harita görüntüsü[5]

B. GAN Mimarisi

GAN mimarisi iki ağ yapısından oluşur: ayırıcı ve üretici. Ayırıcı derin ağ, sentetik olarak üretilen sahte görüntüler ile girdi olarak alınan gerçek görüntüleri birbirinden ayırt etmeye çalışır. Üretici derin ağ ise giriş olarak alınan bir görüntü sinyalinden gerçek görüntülere benzer görüntüler üretmeye çalışır. Ancak üretilen görüntüler gerçekmiş gibi ayırıcıyı kandırmaya çalışır. Üretici rastgele sayıları alır ve bir görüntü döndürür. Oluşturulan bu görüntü, gerçek, temel gerçeklik veri setinden alınan bir görüntü akışıyla birlikte ayırıcıya beslenir. Ayırıcı, hem gerçek hem de sahte görüntüleri alır ve 1'in bir özgünlük tahminini ve 0'ın sahteyi temsil ettiği 0 ile 1 arasında bir sayı olan olasılıkları döndürür.

Ayırıcı ağ eğitim sırasında, hem gerçek verileri hem de üreticiden gelen sahte verileri sınıflandırır. Ayırıcı ağ döndürdüğü olasılık değerleri ile olması gereken değerler arasındaki fark olan kayıp(loss) değeri kullanılarak eğitilir. Ayırıcı ağda geriye yayılım yoluyla ağırlıklar güncellenir.

GAN mimarisi şekil 2'de verilmiştir.

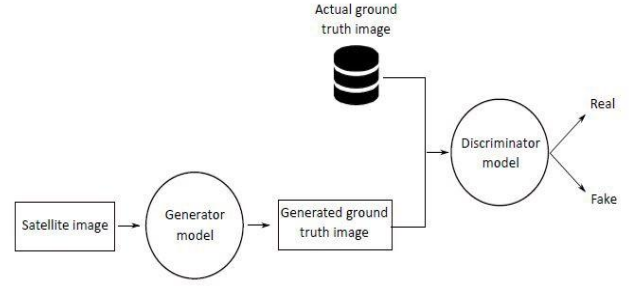


Şekil 2. GAN Mimarisi

Önerdiğimiz yöntemde GAN mimarilerinden biri olan Pix2Pix GAN kullanılmıştır. Pix2Pix, genel amaçlı görüntüden görüntüye çeviri için tasarlanmış bir GAN modelidir. Pix2Pix modeli, çıktı görüntüsünün oluşturulmasının bir kaynak görüntüye bağlı olduğu bir koşullu GAN veya Conditional GAN(cGAN) türüdür[5]. Pix2Pix GAN, haritaları uydu fotoğraflarına, siyah beyaz fotoğrafları renklilere ve ürün eskizlerini ürün fotoğraflarına dönüştürmek gibi bir dizi görüntüden görüntüye çeviri görevlerinde kullanılmıştır.

Üretici, U-Net mimarisi kullanan bir kodlayıcı-kod çözücü modelidir. Girdi olarak kaynak görüntüyü (uydu görüntüsü) alır ve çıktı olarak hedef görüntüyü (google harita görüntüsü) oluşturur. Üreteç modeli, görüntüyü belirli bir düzeye kadar altörnekleyerek kodlar ve U-Net mimarisini kullanarak görüntünün orijinal boyutuna yükselterek kodunu çözer.

Ayrıtıca hem kaynak görüntü hem de hedef görüntü sağlanır ve hedefin kaynak görüntünün uygun bir dönüşümü olup olmadığını belirlenmesi gerekir. Hem kaynak görüntüyü (örn. uydu fotoğrafı) hem de hedef görüntüyü (örn. Google harita görüntüsü) girdi olarak alır ve hedef görüntünün gerçek mi yoksa kaynak görüntünün sahte bir çevirisi mi olduğu olasılığını tahmin eder. Ayrıtı edici tasarımı, modelin bir çıktısı ile girdi görüntüsündeki piksel sayısı arasındaki ilişkiyi tanımlayan modelin etkin alıcı alanını temel alır. Oluşturucu, üreticiyi hedef etki alanında makul görüntüler oluşturmaya teşvik eden çekişmeli kayıp yoluyla eğitilir. Üretici ağ ayrıca, oluşturulan görüntü ile beklenen çıktı görüntüsü arasında ölçülen L1 kaybı aracılığıyla da güncellenir. Bu ek kayıp, üreteç modelini kaynak görüntünün makul çevirilerini oluşturmaya teşvik eder. Önerilen modelin işleyişi şekil 3'te verilmiştir.

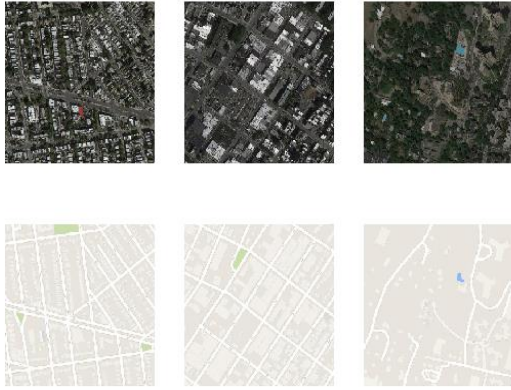


Şekil 3. Önerilen Model

III. BULGULAR

Eğitim veri setindeki 1096 görüntü 100 adımda eğitilmiştir. Üretici her 10 adımda veya 10.960 eğitim adımında bir kaydedilir ve değerlendirilir ve model 100 adım veya toplam 109.600 eğitim adımı boyunca çalışır. Her eğitim adımı, önce bir grup gerçek örnek seçmeyi, ardından gerçek kaynak görüntüleri kullanarak bir grup eşleşen sahte örnek oluşturmak için oluşturucuyu kullanmayı içerir. Ayrıtıca daha sonra gerçek görüntüler ve ardından sahte görüntüler yığını ile güncellenir. Daha sonra, girdi olarak gerçek kaynak görüntüleri sağlayarak ve kayıp hesaplamak için gereken modelin beklenen çıktıları olarak 1 (gerçek) sınıf etiketlerini ve gerçek hedef görüntüleri sağlayarak üreteç modeli güncellenir. Son olarak, her güncelleme için kayıp, her eğitim yinelemesinde rapor edilir ve model performansı her 10 eğitim adımında değerlendirilir.

Sonuç olarak, uydu görüntülerinin google map görüntüsüne geçişini elde etmiş oluruz. Sistem, benzerlik açısından görsel olarak analiz edilebilen belirli bir uydu görüntüsünün, karşılık gelen google harita görüntüsünün ve beklenen/oluşturulan google harita görüntüsünün çıktısını vermektedir. Model test seti üzerinde değerlendirildiğinde elde edilen örnek çıktılar Şekil 4 ve 5'te gösterilmektedir. Şekil 5'te birinci sütun uydu görüntülerini, ikinci ve üçüncü sütun ise verilen uydu görüntüsü için oluşturulan ve gerçek harita görüntülerini içerir.



Şekil 4. Pix2Pix GAN Modeli ile Uydudan Google Haritaya Görüntü Çevirisi



Şekil 5. Pix2Pix GAN Modeli ile Uydudan Google Haritaya Görüntü Çevirisi

IV. SONUÇLAR

Bir uydu görüntüsünü karşılık gelen harita görüntüsüne dönüştüren GAN tabanlı bir model önerdik. Önerilen Pix2Pix GAN modeli ile kaynak uydu görüntülerine son derece benzer görüntüler elde edilmiştir. Model daha önce görmediği veriler üzerinde başarılı sonuçlar üretmektedir. Elde edilen harita görüntülerinin bir kısmında sokakların düz çizgiler gibi görünmemesi ve binaların bazı detaylarının eksik olması gibi sorunlar, eğitim süresinin artırılması veya farklı bir model seçimi ile iyileştirilebilecektir. Böylece daha yüksek kalitede görüntüler oluşturularak ince ayrıntıların kaybolmaması sağlanacaktır.

KAYNAKLAR

[1] Sadeep, J., 2015. Semantic Image Segmentation with Deep Learning.

[2] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, pp. 2672– 2680, 2014.

[3] S. Martin Arjovsky and L. Bottou, "Wasserstein generative adversarial networks," in *Proceedings of the 34 th International Conference on Machine Learning*, Sydney, Australia, 2017.

[4] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.

[5] P. Isola, J. Zhu, T. Z. Alexei, A. Efros, *Image-to-Image Translation with Conditional Adversarial Networks*, Berkeley AI Research (BAIR) Laboratory, UC Berkeley.

[6] J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2242– 2251.

[7] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European conference on computer vision*. Springer, 2016, pp. 694–711.

[8] T. Kim, B. Kim, M. Cha, and J. Kim, "Unsupervised visual attribute transfer with reconfigurable generative adversarial networks," *arXiv preprint arXiv:1707.09798*, 2017.

[9] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.

[10] Deng, X.; Zhu, Y.; and Newsam, S. 2018. What Is It Like Down There? Generating Dense

Ground-Level Views and Image Features From Overhead Imagery Using Conditional Generative Adversarial Networks. *SIGSPATIAL 18*, November 2018, Seattle, Washington USA.

[11] Costea, D.; Marcu, A.; Slusanschi, E.; and Leordeanu, M. 2017. Creating Roadmaps in Aerial Images with Generative Adversarial Networks and Smoothing-based Optimization. *ICCV 2017 Workshops*.

[12] Zhu, J.-Y.; Park, T.; Isola, P.; and Efros, A. A. 2018. Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv:1703.10593*.

[13] Xu, C., and Zhao, B. 2018. Satellite Image Spoofing: Creating Remote Sensing Dataset with Generative Adversarial Networks. *10th International Conference on Geographic Information Science*.