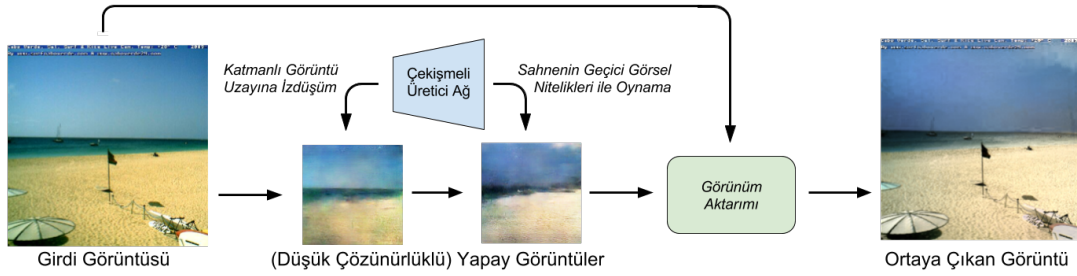


Çekişmeli Üretici Ağlar Kullanarak Dış Mekan Görüntülerinin Geçici Niteliklerini Düzenleme

Adjusting Transient Attributes of Outdoor Images using Generative Adversarial Networks

Levent Karacan, Aykut Erdem, Erkut Erdem
Hacettepe Üniversitesi Bilgisayarlı Görü Laboratuvarı, Ankara, Türkiye
{karacan, aykut, erkut}@cs.hacettepe.edu.tr



Şekil 1: Bir dış mekan görüntüsü verildiğinde yaklaşımımız bu görüntünün farklı zaman (gün batımı, gece, vb.) veya farklı doğa koşullarındaki (sisli, puslu, vb.) bir halini otomatik olarak oluşturabilmektedir. Örnek görüntünün “yağmur” niteliği artırılmıştır.

Özetçe —Sürekli değişim halinde bulunan dünyamızda, doğal açık hava sahneleri günün ve yılın farklı zamanlarında aydınlatma, hava durumu ve mevsim koşullarından dolayı önemli değişiklikler yaşarlar. Bu nedenle, dış mekan görüntülerinin görünümünü gerçekçi biçimde otomatik olarak değiştirebilen hesaplamalı modelleri oluşturmak oldukça zorludur. Önerilen yaklaşımlar, büyük bir veritabanından benzer görüntüleri getirmek ve bu görüntüleri girdi görüntüsü ile eşleştirmek gibi nihai sonucu ciddi olarak etkileyebilen çeşitli ara adımlar kullanmaktadır. Önceki yöntemlerin bu dezavantajlarını ortadan kaldırmak yönünde yapılan bir çaba olarak, bu çalışmada dış mekan görüntülerinin katmanlı görünüm uzayını çekişmeli üretici ağlar kullanarak öğrenen bir otomatik görüntü düzenleme yaklaşımı sunulmaktadır. Yaptığımız deneyler, modelimizin doğal duran umut verici sonuçlar verdiğini göstermektedir.

Anahtar Kelimeler—hesaplamalı görüntüleme, sahne analizi, görüntü düzenleme, üretici çekişmeli ağlar, geçici nitelikler.

Abstract—In our ever changing world, natural outdoor scenes undergo significant changes due to lighting, weather and seasonal conditions at different times of the day and the year. Therefore, it is remarkably challenging to build computational models which can automatically manipulate the appearance of outdoor images in a realistic manner. Suggested approaches employ several intermediate steps that may seriously affect the quality of the result, such as retrieving similar images in a large database and matching those images to the input image. As an effort to eliminate these drawbacks of the previous methods, in this paper, we present an automatic image editing approach which utilizes generative adversarial networks to learn the appearance manifold of outdoor images. Our experiments show that our model yields natural looking promising results.

Keywords—computational photography, scene analysis, image editing, generative adversarial networks, transient attributes.

I. GİRİŞ

Açık hava sahnelerinin görünüşleri, günün ve yılın farklı zamanlarında aydınlatma, hava durumu ve mevsim koşullarında ortaya çıkan farklılıklardan ötürü çok büyük değişiklikler gösterebilmektedir. Fotoğrafçılık açısından düşüncecek olursak fotoğrafçılar daha çekici görüntüler elde edebilmek için özellikle gün doğumu veya gün batımı zamanlarını tercih ederler ama böyle bir gün batımı etkisini bir görüntüye otomatik yollarla ve doğallığı bozmadan vermek o kadar kolay değildir. Dış mekan sahnelerinin bu tarz geçici görsel niteliklerinin (gece, günbatımı, sis, bulut, vb.) hesaplamalı yöntemler ile düzenlenmesi bu bakımdan ilgi çekici bir problemdir. Matematiksel olarak bu problemin çözümü aslen dış mekan görüntülerinin geçici niteliklerini de kodlayabilen katmanlı bir görüntü uzayının (*image manifold*) modellenmesini ve öğrenilmesini gerektirmektedir.

Geçmişte geçici görsel niteliklerin düzenlenmesi için yapılan çalışmalar detaylarını bir sonraki bölümde aktaracağımız veriye dayalı yaklaşımları kullanmışlardır [1], [2]. Shih *vd.*, zamana bağlı (*time-lapse*) videolar üzerinden günün farklı zamanlarında dış mekan görüntülerinde meydana gelen değişimi düzenlemeyi amaçlarken [1] bir başka çalışmada Laffont *vd.* ise mevsimsel etkilerden kaynaklanan daha genel değişimleri düzenleme üzerine gitmişler ve bu amaçla geçici görsel nitelikler ile etiketlenen yeni bir veri kümesi oluşturmuşlardır [2]. Bu çalışmamızda ilgili problem, öğrenmeye dayalı bir yaklaşım üzerinden ele alınmakta ve bu amaçla çeşitliliği yüksek bir takım açık hava görüntüsü üzerinde geçici görsel niteliklere koşullu olarak eğitilen bir çekişmeli üretici ağ (*generative adversarial network* – GAN) [3] modelinin kullanım yolu araştırılmaktadır. Şekil 1’de özetlendiği üzere verilen bir girdi

görüntüsü öncelikle GAN modelinin öğrendiği katmanlı görüntü uzayına geri düşürülmekte ve bu sayede üretilen sahnenin nitelikleri üzerinde kolaylıkla oynanabilmektedir. Son adımda ise yüksek çözünürlü ve detaylı bir sonuç için görünüm aktarımı yapılmaktadır. Böylelikle önceki çalışmalardan farklı olarak benzer görüntü bulma ve bulunan görüntü ile girdi görüntüsü arasında başarılı bir görünüm aktarımı yapılabilmesi için eşleme yapılması adımına gerek kalmamaktadır.

II. İLGİLİ ÇALIŞMALAR

A. Görüntü Düzenleme

Görüntülerin görünümünü düzenlemek üzerinde uzun zamandır çalışılan bir uygulama alanıdır. Bir girdi görüntüsünün bütünsel renk bilgisini kaynak bir başka görüntünün renk tonlarına benzetmeyi amaçlayan çalışmalar algısal renk uzayında basit istatistiksel yöntemleri kullanırlar [4], [5]. Levin ve Weiss'in öncül renklendirme çalışmasında ise kaynak görüntü yerine girdi olarak kullanıcılardan yerel renk bilgisi sağlamaları beklenmekte ve bu renkler ilgili görüntüye eniyileme yoluyla yayılmaktadır [6]. Bu tarz renk bilgisi düzenleme üzerine giden yaklaşımlara ek olarak kullanıcı kontrollü görüntü düzenleme literatüründe yapısal bilgilerin de düzenlendiği bazı çalışmalar mevcuttur [7].

Geçici görsel niteliklerin düzenlenmesini amaçlayan çalışmalar, dış sahnelerdeki zamansal ve mevsimsel değişimleri modelleyebilmek için aynı sahnenin farklı zaman veya mevsimde çekilmiş görüntülerini kullanarak bu geçişleri modellemeye çalışmışlardır [1], [2]. Shih *vd.* verilen bir girdi görüntüsü ve bir hedef zaman için zaman aralıklı video verilerini kullanarak girdi görüntüsüne hedef zamanda olması beklenen geçici görsel niteliği kazandırmayı amaçlamıştır [1]. Daha geniş geçici görsel niteliklerin de düzenlenmesi için Laffont *vd.* mevsimsel değişiklikleri de barındıran bir veri kümesi oluşturmuş ve verilen bir görüntünün geçici görsel niteliklerini tahmin etmek için bağlanım (*regression*) analizi yapmayı önermişlerdir. Bu şekilde kullanıcı girdileriyle geçici nitelik uzayında istenilen nitelikteki görüntü rahatça bulunabilmekte ve girdi görüntüsüne görünüm aktarımı (*appearance transfer*) gerçekleştirilmektedir [2]. Bu çalışmalar, belli bir seviyeye kadar başarılı olsalar da istenilen sonucun elde edilmesi aranan nitelikteki görüntünün veri kümesinde bulunmasına bağlıdır. Ayrıca bulunan görüntünün aktarımının yapılacağı girdi görüntüsüne içerik olarak belli oranda benzemesi de gerekmektedir.

B. Derin Üretici Modeller

Yakın geçmişte görüntülerin yaşadığı katmanlı uzayları başarılı bir şekilde modelleyebilen derin üretken modeller literatürde önerilmeye başlanmıştır [3], [8]. Bu çalışmaların devamında gelen çalışmalar bazı öğrenme kararsızlıklarını çözmüş ve daha kaliteli görüntülerin üretilmesini sağlamıştır [9]–[12]. Öte yandan sınıf etiketleri [9], [13], görüntü altıyazıları [14], nesneleri kapsayan kutular ve nesnelere ait anahtar noktalar [15] gibi ek bilgiler de görüntü üretme işleminde koşul olarak kullanılarak görüntü üretme işlemini kontrol edilebilen bazı çalışmalarda ortaya çıkmıştır. Koşul olarak kullanılan bu ek bilgileri ve yeni teknikler verilen bir veri kümesinin modellenmesini ve yeni örneklerin kontrollü bir şekilde katmanlı uzay üzerinden doğal bir biçimde üretilmesini sağlamıştır.

Özellikle katmanlı uzay üzerinden üretilen görüntülerin doğal ve yumuşak geçişlere olanak tanınması birçok araştırma-

cının dikkatini çekmiştir ve önemli uygulama alanları ortaya çıkmıştır. [16]'da yazarlar evrimsel sinir ağları kullanarak özgün 3 boyutlu sandalye modelleri üretmeyi öğrenen bir model geliştirmişler ve bu modeli mevcut sandalyelerin biçimlerinin düzenlenmesinde kullanmışlardır. Bir başka çalışmada, yüz görüntüleri için nitelik koşullu modeller kullanılarak çeşitli yüz niteliklerinin düzenlenmesi başarılı bir şekilde gerçekleştirilmektedir [17]. Çalışmamızla yakından ilişkili olan [18]'de ise; yazarlar koşulsuz çekışmeli bir üretici derin model [9] kullanarak katmanlı bir görüntü uzayı öğrenmekte ve uzay üzerindeki temsile dayalı bir düzgünleştirme teriminden yararlanarak bir görüntü düzenleme yaklaşımı önermektedirler. Böylelikle kullanıcıların istekleri/girdileri çerçevesinde görüntülere çeşitli görünüm aktarımları yapılabilmektedir ve veri kümesinde olmayan yeni örnekler üretilabilmektedir.

III. ÖNERİLEN YAKLAŞIM

A. Çekişmeli Üretici Ağlar

Çekişmeli Üretici Ağlar, iki oyunculu bir minimum-maksimum oyunu içinde birbirleriyle yarışan; G ile ifade edilen bir üretici ağdan ve D ile ifade edilen bir ayırt edici ağdan oluşmaktadır [3]. Bu yapıda hem üretici, hem de ayırt edici ağ çok katmanlı algılayıcılar kullanılarak modellenmektedir. Ayırt edici ağın gerçek görüntüleri sentetik olarak üretilen görüntülerden başarılı bir şekilde ayırmayı amaçlamakta iken üretici ağ ise gerçek görüntülere olabildiğince benzeyen yapay görüntüler üretmek bu ayırt edici ağı aldatmaya çalışmaktadır. Bu oyun aşağıdaki eniyileme problemi şeklinde modellenerek üretici ve ayırt edici ağların bir arada eğitilmesi sağlanmaktadır:

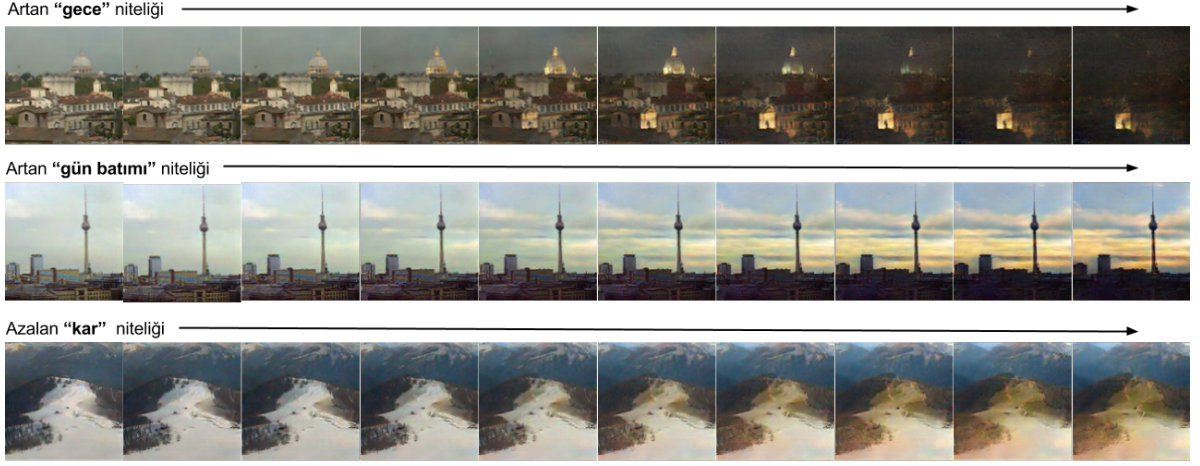
$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{x \sim p_z(z)} [\log (1 - D(G(z)))] \quad (1)$$

Burada x gerçek veri dağılımı $p_{data}(x)$ 'ten gelen doğal bir görüntüyü ifade ederken; z de tekdüze bir dağılımdan örneklenen rassal saklı bir vektörü temsil etmektedir. Yeterli sayıda eğitim görüntüsü ile ve belli sayıda gerçekleşen, üretici ve ayırt edici ağ üzerinde birbirini izleyen model güncellemeleri sonucunda üretici model G 'nin ürettiği görüntülerin dağılımı, gerçek veri dağılımı $p_{data}(x)$ 'ye yakınsayabilmektedir [3]. Bu sayede, üretici ağ bir rassal vektör z 'den gerçek görüntülerden rahatlıkla ayırt edilemeyecek sentetik bir görüntü $G(z)$ üretebilmektedir.

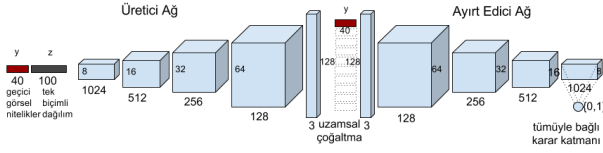
B. Sahne Niteliğine Bağlı Görüntü Üretimi

Çekişmeli üretici ağların genişletilmiş bir hali olan şartlı çekişmeli üretici ağlar [9], [10], görüntü üretimi için katmanlı görüntü uzayını kodlayan rassal vektör z 'nin yanında bazı ek bilgileri de kullanmaktadır. Formal olarak genelde izlenen yol, ek bilgilerin c ile ifade edilen bir vektörle temsil edilmesi ve bu vektörün rassal vektör z 'e ardarda bağlanması ile elde edilen bütünselik vektörle üretici ve ayırt edici ağların beslenmesidir. Bu sayede $G(z, c)$ ile gösterilen genişletilmiş üretici ağ, gerçekçi bir görüntü üretirken de; $D(x, c)$ ile ifade edilen genişletilmiş ayırt edici ağ üretilmiş ve gerçek görüntüleri ayırt ederken de ilgili işlemler bağlam vektörü c 'nin kontrollü altında gerçekleşmektedir.

Şekil 2'de geçici görsel nitelikleri önkoşul olarak kullanıldığımız evrimsel sinir ağı mimarisi gösterilmektedir. Görülebildiği gibi ağ, üretici ve ayırt edici olmak üzere iki kısımdan oluşmaktadır. Üretici ağda 100 boyutlu bir örnek



Şekil 3: Bazı geçici görsel nitelikler için öğrenilmiş olan nitelik koşullu katmanlı uzaydan üretilen görüntüler.



Şekil 2: Çalışmamızda kullandığımız geçici görsel niteliklerin ön koşul olarak kullanıldığı üretici çekişmeli ağ mimarisi.

dağılımdan elde edilen rassal z ve önkoşul olarak kullanılan 40 boyutlu geçici görsel nitelikler ardarda bağlandıktan sonra öznetelik eşlemlerine (*feature maps*) dönüştürülmekte ve sonrasında peşi sıra 4 ters evrişim katmanı ile 128×128 boyutunda renkli görüntü üretimi gerçekleştirilmektedir. Ayırt edici ağda 40 boyutlu geçici görsel nitelikler önkoşul olması için uzamsal olarak çoğaltılarak olarak 128×128 boyutlu renkli görüntüye bağlanmaktadır. Sonrasında peşisıra 4 evrişim katmanı sonrasında elde edilen öznetelik eşlemleri ağı giren görüntünün verilen önkoşul altında gerçek bir görüntü olup olmadığına karar verilmesi için tamamen bağlı bir katmandan geçirilmektedir. Şekil 3'te öğrenilmiş çok katmanlı görüntü uzayı üzerinde farklı geçici nitelikler kullanılarak yaratılan görüntü örnekleri verilmiştir. Gözükteği üzere öğrenilen görüntü modeli başarılı bir şekilde farklı nitelikler arasında yumuşak bir geçişle görüntü üretebilmektedir.

C. Niteliğe Bağlı Görüntü Düzenleme

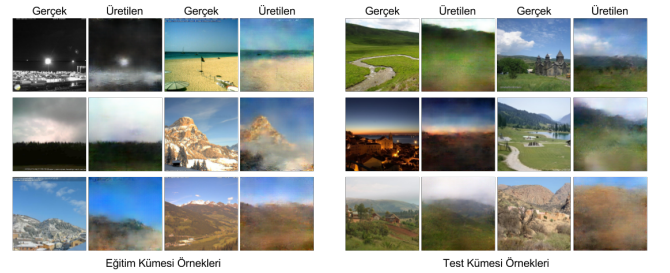
Niteliği değiştirilmek istenen yüksek çözünürlüklü bir x girdi görüntüsü için ilk olarak gerçekleştirilen işlem, bu görüntünün çekişmeli üretici ağın öğrendiği katmanlı görüntü uzayındaki izdüşümünün hesaplanmasıdır. Bunun için öncelikle girdi görüntüsü x üzerinde bir yeniden boyutlandırma gerçekleştirilmekte ve girdi görüntüsünün üretici ağın ürettiği görüntülerle aynı çözünürlüğe indirgenmesi sağlanmaktadır. Boyutu indirgenmiş x_ℓ görüntüsünün katmanlı görüntü uzayındaki izdüşümü, x_ℓ görüntüsünü üreten rassal vektör z^* ve ilgili bağlam vektörü c^* 'nin hesaplanması ile bulunmaktadır. Bu işlem için aşağıdaki eniyileme problemi çözülmektedir:

$$(z^*, c^*) = \operatorname{argmin} \mathcal{L}(g(z, c), x_\ell) \quad (2)$$

Burada $\mathcal{L}$ iki görüntü arasındaki farkı ölçen herhangi bir kayıp fonksiyonu olabilir. Deneylerimizde iki görüntü arasındaki

Öklid kayıp fonksiyonu bu amaçla kullanılmıştır.

Şekil 4'te bu yöntemle elde edilen geriçatma sonuçlarına yer verilmektedir. Görülebileceği üzere, yüksek çözünürlüklü görüntüler ile bu görüntülerin üretici ağın öğrendiği katmanlı görüntü uzayındaki izdüşümleri kullanılarak üretilmiş halleri birbirlerine kabaca benzemektedir. Ancak bazı sahneler için ilgili sahnenin bazı altbileşenleri tamamen yakalanamayabilmektedir. Örneğin, birinci sıradaki kumsal görüntüsünün geriçatma sonucu orjinal görüntüdeki kuma gömülü bayrağı üretememiştir.



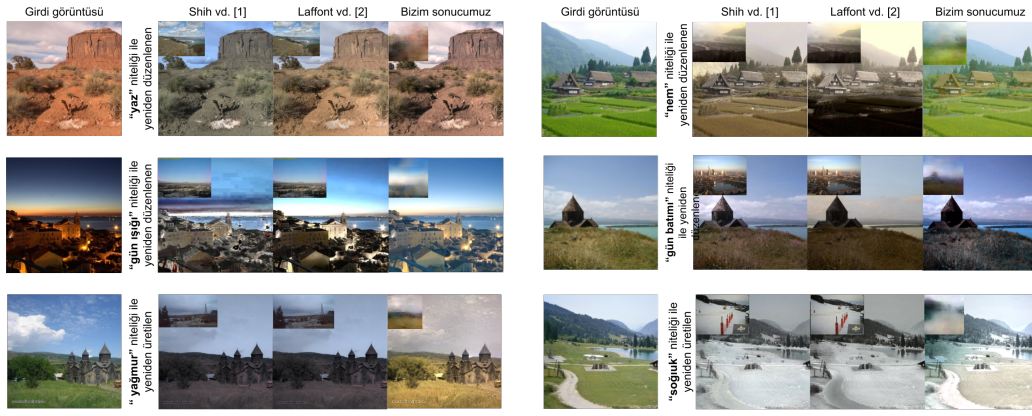
Şekil 4: Bazı görüntüler için elde edilen geriçatma sonuçları.

Çekişmeli üretici modeller oldukça başarılı sonuçlar vermesine rağmen henüz istenilen çözünürlükte görüntü üretimine olanak tanımamaktadır. Bu nedenle daha kaliteli ve yüksek çözünürlükte görüntü edebilmek için girdi görüntüsü ile benzer yapıya sahip yeni nitelikle üretilen görüntüden girdi görüntüsüne görünüm aktarımı yapılmaktadır. Bunun için Shih vd. tarafından önerilen görüntü aktarım yöntemi [1] kullanılmıştır. Bu yaklaşımda kaynak görüntü ile hedef görüntü arasında örnek tabanlı yerel olarak afin renk aktarımı en küçük kareler eniyileme yöntemi ile bulunan afin modellerle gerçekleştirilmektedir.

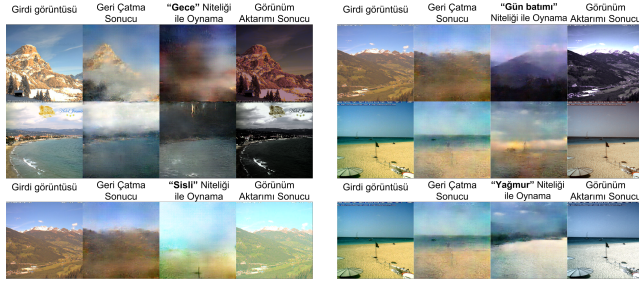
IV. DENEYSEL ANALİZ

A. Veri kümesi ve Önışleme

Transient Attributes veri kümesi [2], dünyanın farklı bölgelerine yerleştirilmiş 101 web kamerasından farklı zaman, tarih ve iklim koşullarında toplanmış 8.571 görüntü içermektedir. Her bir web kamerası birbirlerine hizalanmış ve çeşitliliği



Şekil 6: İlgili çalışmalar ile karşılaştırma. Önerilen model mevcut çalışmalara kıyasla daha başarılı sonuçlar üretmektedir.



Şekil 5: Bazı görüntüler için nitelik düzenleme sonuçları.

yüksek 60 ila 120 görüntü barındırmakta ve her görüntü, gün batımı, karlı, sisli, yağmurlu, vb. önceden belirlenmiş 40 farklı geçici görsel niteliklere göre etiketlenmiştir.

B. Deneysel Sonuçlar

Önerilen yöntem ile çeşitli test görüntüleri için geçici görsel nitelik düzenleme sonuçları Şekil 5'te verilmektedir. Sonuçlarda görüldüğü gibi yeniden üretme sonuçları başarılı olmasına rağmen tahmin edilen nitelikler üzerinde yapılan değişiklikler kısıtlı kalmaktadır. Buna rağmen oldukça göze hoş gelen sonuçlar elde edilmektedir. Ayrıca bazı örnekler için kullanılan görünüm aktarım yöntemi başarısız olmuştur. Örneğin yağmurlu niteliği ile yeniden üretilen görüntü istenilen etkiyi başarılı bir şekilde gösterse de, aktarım yapıldıktan sonra bulut görünümü kaybolmaktadır. Şekil 6'da veriye dayalı yöntemleri kullanan çalışmalara önerilen yöntemin sonuçları karşılaştırılmaktadır. Test görüntüleri eğitimde kullanılan hiçbir web kamerasına ait olmamasına rağmen; bu görüntüler başarılı bir şekilde yeniden üretilmiş ve bu görüntülerin nitelikleriyle oynanarak mevcut yöntemlere kıyasla daha başarılı düzenleme sonuçları elde edilmiştir.

V. SONUÇ VE TARTIŞMA

Önerilen yöntem ile geçici görsel niteliklerin düzenlenmesi için iyi tanımlanmış bir yöntem önerilmiş ve ümit verici sonuçlar alınmıştır. Sonraki çalışmalarda geçici görsel nitelikleri daha doğru tahmin etmek ve düzenlemek için öğrenilen çok katmanlı uzaya izdüşürme yönteminin geliştirilmesi ve renk bilgisinin yanında doku bilgisini de aktarabilen daha başarılı bir görünüm aktarma yöntemi tanımlanması planlanmaktadır.

TEŞEKKÜR

Grafik kartı desteği için NVIDIA'ya teşekkür ederiz.

KAYNAKLAR

- [1] Y. Shih, S. Paris, F. Durand, and W. Freeman, "Data-driven hallucination of different times of day from a single outdoor photo," *ACM Transactions on Graphics*, vol. 32, no. 6, 2013.
- [2] P.-Y. Laffont, Z. Ren, X. Tao, C. Qian, and J. Hays, "Transient attributes for high-level understanding and editing of outdoor scenes," *ACM Transactions on Graphics*, vol. 33, no. 4, 2014.
- [3] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *NIPS*, 2014.
- [4] E. Reinhard, M. Adhikhmin, M. Gooch, and P. Shirley, "Color transfer between images," *IEEE Computer Graphics and Applications*, vol. 21, no. 5, pp. 34–41, 2001.
- [5] F. Pitie, A. C. Kokaram, and R. Dahyot, "N-dimensional probability density function transfer and its application to color transfer," in *ICCV*, 2001.
- [6] D. L. A. Levin and Y. Weiss, "Colorization using optimization," *ACM Transactions on Graphics (ToG)*, vol. 23, no. 3, pp. 689–694, 2004.
- [7] C. Barnes, E. Shechtman, A. Finkelstein, and D. Goldman, "Patchmatch: A randomized correspondence algorithm for structural image editing," *ACM Transactions on Graphics*, vol. 28, no. 3, p. 24, 2009.
- [8] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," in *ICLR*, 2014.
- [9] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," in *ICLR*, 2016.
- [10] E. Denton, S. Chintala, A. Szlam, and R. Fergus, "Deep generative image models using a laplacian pyramid of adversarial networks," in *NIPS*, 2015.
- [11] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training GANs," in *NIPS*, 2016.
- [12] K. Gregor, I. Danihelka, A. Graves, D. Rezende, and D. Wierstra, "Draw: A recurrent neural network for image generation," in *ICML. JMLR Workshop and Conference Proceedings*, 2015.
- [13] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv:1411.1784*, 2014.
- [14] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee, "Generative adversarial text to image synthesis," in *ICML*, 2016.
- [15] S. Reed, Z. Akata, S. Mohan, S. Tenka, B. Schiele, and H. Lee, "Learning what and where to draw," in *NIPS*, 2016.
- [16] A. Dosovitskiy, J. T. Springenberg, and T. Brox, "Learning to generate chairs with convolutional neural networks," in *CVPR*, 2015.
- [17] X. Yan, J. Yang, K. Sohn, and H. Lee, "Attribute2image: Conditional image generation from visual attributes," in *ECCV*, 2016.
- [18] J.-Y. Zhu, P. Krähenbühl, E. Shechtman, and A. A. Efros, "Generative visual manipulation on the natural image manifold," in *ECCV*. Springer, 2016, pp. 597–613.