

Derin Öğrenme ile Zemin Dokusu Sınıflandırma

Ground Texture Classification with Deep Learning

Mustafa Özuysal
Bilgisayar Mühendisliği Bölümü
İzmir Yüksek Teknoloji Enstitüsü
mustafaouysal@iyte.edu.tr

Özetçe —Bu çalışmada ImageNet veri setinde daha önceden eğitilmiş farklı mimarideki derin sinir ağlarının transfer öğrenmesi yolu ile zemin dokularının sınıflandırılması için kullanılması araştırılmıştır. Yedi farklı zeminden toplanan görüntüler ile yeni bir zemin dokusu veri seti oluşturulmuştur. Bu veri seti ile derin sinir ağları kısmen ya da mümkün olduğunda tüm katmanlarıyla yeniden eğitilmiştir. Sonuçlar küçük imgeler kullanıldığında bile zemin dokularının başarıyla sınıflandırıldığını göstermektedir.

Anahtar Kelimeler—Doku sınıflandırma, derin sinir ağları, transfer öğrenmesi

Abstract—In this study, we investigate the use of transfer learning on various deep neural network architectures pretrained on the ImageNet data set for ground texture classification purposes. We introduce a new ground texture data set collected from seven different areas. We retrain deep neural network's last layer or when possible the full set of layers on this data set. The results show that it is possible to discriminate the ground textures even when very small images are used.

Keywords—Texture classification, deep neural networks, transfer learning

I. GİRİŞ

Zemin görüntüleri hem hangi mekanda bulunduğu hem de bu mekanda yer alması olası nesnelere hakkında bilgi içerir. Özellikle iç ve dış mekan imgelerinde zeminde yer alan dokular farklılık gösterir. Örneğin ahşap bir zemin genellikle iç mekanda bulunduğu işaret eder. Ayrıca bir mekanda bulunan zemin tipi, zemin üzerinde yer alabilecek nesnelere de etkilemektedir. Bu nedenle zemin dokularının tanınması konumlandırma ve haritalama uygulamalarında yardımcı bir öğe olarak kullanılabilir.

Doku sınıflandırma, farklı desen ve istatistik özelliklere sahip imgelerin birbirinden ayrılmasını gerektirir. Bu nedenle doku sınıflandırma yöntemleri imgeler üzerinde birden fazla filtrenin sonuçlarını bir arada kullanırlar. Geçmişte Gabor filtreleri gibi sabit filtre grupları [1], [2] kullanılırken, son zamanlarda filtreleri de makine öğrenme probleminin parçası haline getiren derin öğrenme tabanlı yaklaşımlar da geliştirilmiştir [3]–[5].

Bu bildiride yer alan çalışmalar İzmir Yüksek Teknoloji Enstitüsü tarafından 2017IYTE69 numaralı bilimsel araştırma projesi kapsamında desteklenmiştir.

Bu çalışmada farklı iç ve dış mekanlardan toplanan zemin görüntülerinin sınıflandırılması amacıyla ImageNet [6] veri setinde eğitilmiş farklı mimarideki derin sinir ağları kullanılmıştır. ImageNet üzerinde eğitilmiş sinir ağları başlangıç noktası olarak alınmış ve zemin görüntülerinin tanınması için transfer öğrenmesi [7] yöntemiyle tekrar eğitilmişlerdir. Burada amaç her bir zemin için görece az sayıda imge ile eğitimin sağlanabilmesidir. ImageNet veri setinde yer alan imgeler ile zemin doku imgeleri benzerlik içerdiğinden öğrenilen evrişimli katman filtrelerinin de zemin dokusu sınıflandırılmasında başarı göstereceği beklenebilir.

Çalışma sırasında birbirinden farklı mimaride derin sinir ağları kullanılarak bunların zemin dokusu sınıflandırma amacına uygunlukları karşılaştırılmıştır. Farklı mimariler transfer öğrenmesine farklı derece uygunluk göstermekte ve bazı mimariler daha başarılı sonuç elde etmektedir. Kullanılan veri seti görece küçük olduğundan transfer öğrenmesiyle bazı mimarilerin tüm katmanlarının yeniden eğitilmesi mümkün olmamıştır. Genelde sinir ağının sadece son katmanları eğitilerek evrişimli filtrelerden oluşan alt katmanlar sabit bir filtre grubu olarak kullanılmıştır. Ancak veri yeterli olduğu durumlarda bu katmanların da eğitilmesinin başarıyı artırdığı gözlenmiştir.

Derin sinir ağı mimarisi yanında imge boyutunun sınıflandırmaya etkisi ölçülen bir başka değişkendir. Ayrıca imge boyutları hem eğitim hızını hem de kullanılması gereken ekran kartı hafıza miktarını etkilemektedir. Sonuçlar küçük imgeler kullanıldığında bile farklı zemin dokularının birbirinden ayrılabilirliğini göstermektedir.

II. İLGİLİ LİTERATÜR

Doku sınıflandırma üzerine literatürde pek çok farklı yaklaşım mevcuttur. Özellikle tekrarlanan motiflerden oluşan ya da istatistiksel olarak sabitlik gösteren dokuların tanınması için seçilen bazı özniteliklerin imge üzerindeki dağılımları sıkça kullanılmaktadır. Sınıflandırma için kullanılacak öznitelikler uygulamaya bağlı olarak Gabor [8] ya da Dalgacık filtreleri [9], Yerel İkili Örüntüler (LBP) öznitelikleri [1] veya sadece parlaklık istatistikleri [10] olmuştur. Öznitelikler hesaplandıktan sonra dağılım istatistikleri çıkarılıp doğrudan ya da kümelenecek sınıflandırma için kullanılmıştır. Dağılım istatistikleri normalize edilerek ya da Fisher yöneylerine dönüştürülerek de kullanılmaktadır. Elde edilen nihai öznitelikler en yakın K-komşu ya da destek yöney makineleri benzeri farklı sınıflandırıcılar yardımıyla kullanılabilir.

Ancak son zamanda geliştirilen yaklaşımlarda sabit öz-nitelikler yerine kullanılacak filtrelerin de veriden öğrenilmesi ön plana çıkmıştır. Evrişimli katmanlarıyla derin sinir ağı, hesaplanacak öz-niteliklerin doğrudan öğrenilebilmesini sağlamaktadır. Cimpoi ve diğerleri [3] ImageNet üzerinde eğitilmiş VGGNet [11] sinir ağının evrişimli sinir ağı katmanlarını alarak, bu katmanları doku imgelerinin sınıflandırılması için kullanmıştır. Evrişimli katmanlar doku imgelerinde tekrar eğitilmemiş, bunun yerine imge üzerindeki tepkileri bir Fisher yöneyinde toplanarak sınıflandırma için kullanılmıştır. Buna alternatif olarak [4] doku imgelerinin tanınması için özel olarak tasarlanmış Doku Evrişimli Sinir Ağı (T-CNN) mimarisini önermiştir. Bu mimari görece kısa ve az sayıda parametre içermektedir ve evrişimli katmanların da sıfırdan eğitilmesini gerektirir.

Bu çalışmada ise bu iki farklı yaklaşımın bir bileşimi olan bir yöntem denenmiştir. ImageNet veri kümesinde eğitilmiş farklı mimarideki derin sinir ağlarının evrişimli katmanları sabit tutularak ya da tümüyle eğitilerek zemin dokularının sınıflandırılması için kullanılmıştır. Veri setinin büyüklüğü de göz önünde bulundurularak görece daha az parametre içeren mimariler tercih edilmiştir. Bu hem [11] tarafından önerilen yöntem aksine ayrıca Fisher vektörlerinin çıkarılıp öğrenilmesini gerektirmez hem de [4] tarafından önerilen sıfırdan eğitim yapılmasına göre daha basittir. Sonuçlar doğru mimari ve eğitim yöntemi seçildiğinde zemin dokularını başarıyla sınıflandıran bir sinir ağının kolaylıkla eğitilebildiğini göstermektedir.

III. YÖNTEM VE DENEYLER

Zemin dokusu sınıflandırılması için derin sinir ağlarını eğitmek amacıyla bir veri seti toplanmıştır. Bu veri setinden örnek görüntüler Şekil 1’de verilmiştir. Toplam yedi farklı mekanda bakış açısı ve ışıklandırma değiştirilerek örnek imgeler alınmıştır. Üç sınıf iç mekan görüntülerinden, diğer dört sınıf ise dış mekan görüntülerinden oluşmaktadır. Bazı görüntülerde doku rengi değişmekte, dış mekanlarda su birikintileri, bitki ve ot kümeleri şaşırtıcı etmen olarak yer almaktadır. Yedi sınıfın her birinde de yüzer adet görüntü bulunmaktadır. Bunlardan 70 tanesi rastgele olarak eğitim kümesi için seçilmiş, kalanlar ise doğrulama ve test kümeleri için ayrılmıştır.

Bu veri kümesiyle çalışmak üzere farklı mimarilerde ve ImageNet üzerinde eğitilmiş derin sinir ağları kullanılmıştır. Mimari olarak ResNet [12], DenseNet [13] ve ResNext [14] seçilmiştir. ResNet mimarisinde evrişimli blok çıktıları, etkisiz bir bağlantı ile birleştirilir. ResNext mimari olarak ResNet’e çok yakındır ancak aynı anda hesaplanan evrişimli blok çıktıları birleştirerek farklı bir blok yapısı elde edilmiştir. DenseNet ise evrişimli blokların ResNet’e göre daha da geriden gelen etkisiz bağlantılar ile çalışması üzerinedir. Önceden eğitilmiş ResNet ve DenseNet modelleri için PyTorch kütüphanesinden [15], ResNext modeli için ise fast.ai kütüphanesinden [16] faydalanılmıştır.

Seçilen bu mimarilerin ortak yanı en son katmanlarının tamamen-bağlı bir sinir ağı yapısı içermesi, bunun dışındaki katmanların ise evrişimli bloklardan oluşmasıdır. Bu sayede, farklı uygulama alanları için farklı imge boyutları bu mimarilere beslenebilir ve yine farklı sayıda sınıf için tamamen-bağlı bir son katman eğitilebilir.



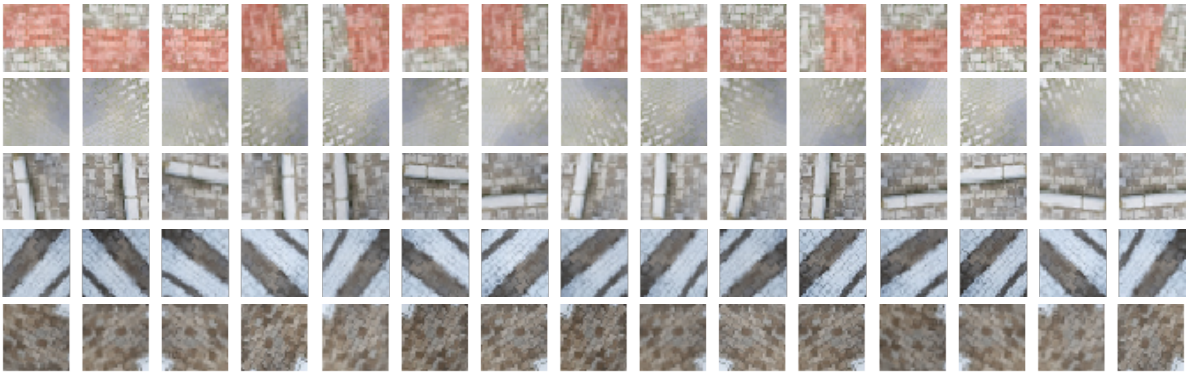
Şekil 1: Zemin dokusu veri setindeki sınıflardan örnekler.

Seçilen mimarilerin eğitim ve sınıflandırma zamanını kısaltmak için bu çalışmada 32×32 ve 64×64 boyutlarında görece düşük boyutlu imgeler ile çalışılmıştır. Tüm mimariler için farklı sayıda temel blok birleştirilerek daha derin ya da sığ sinir ağları eğitmek mümkündür. Yine hesaplama maliyetini azaltmak için daha az sayıda parametre içeren ResNet34, ResNet50, DenseNet121 ve ResNext50 sinir ağlarıyla çalışılmıştır. Deneylerde hem evrişimli katmanlar sabit tutularak son katmanın eğitimi denenmiş hem de tüm katmanların hazırlanan veri setinde eğitilmesi denenmiştir. Tüm aşamalar için kullanılan grafik kartı hafızası yaklaşık 1 GB’dır. Bir seferde işlenen eğitim verisi sayısı olan yığın boyutu genelde 64’tür. Ancak DenseNet121’in tüm katman eğitimleri sırasında hafıza kullanımını diğerlerine benzer tutmak amacıyla yığın boyutu 32’ye düşürülmüştür.

İmgelere eğitim sırasında çeşitlilik katmak ve farklı bakış açılarından veri sayısını artırmak için kısıtlı ölçekleme ve dönme dönüşümleri uygulanmıştır. Bu sayede veriler eğitim sırasında her kullanıldığında farklı görüntüler ile eğitim yapılabilmesi sağlanmıştır. Şekil 2’de uygulanan dönüşümlerden örnekler yer almaktadır.

Eğitim sırasında öncelikle son katmanların eğitimi yapılmıştır. Burada ImageNet veri seti üzerinde eğitilen evrişimli katmanlar sabit birer öz-nitelik hesaplayıcı olarak görev yapmaktadır. Son katman eğitildikten sonra alt evrişimli katmanlar da serbest bırakılarak tüm sinir ağının eğitimi yapılmaya çalışılmıştır. Ancak veri sayısı az olduğu için alt katmanlar daha düşük bir öğrenme hızı ile eğitilmiştir. Eğitim sırasında eğitim hızı dalgalandırılarak parametre uzayında kayıp yüzeyinin uygun olmayan noktalarından uzaklaşmaya çalışılmıştır [17], [18].

32×32 imgeler için elde edilen sonuçlar Tablo I’de veril-



Şekil 2: 32×32 boyutundaki imgelere uygulanan dönüşümlerden örnekler.

tex_000	13	0	1	0	1	0	0
tex_001	0	14	0	0	0	1	0
tex_002	0	0	15	0	0	0	0
tex_003	0	0	0	15	0	0	0
tex_004	0	0	0	0	15	0	0
tex_005	0	0	0	0	0	15	0
tex_006	0	0	0	0	0	0	15
	tex_000	tex_001	tex_002	tex_003	tex_004	tex_005	tex_006

Şekil 3: 32×32 imgeler için ResNeXt50 karışıklık matrisi

miştir. İmge sayısının azlığı ve imge boyutunun küçük olması nedeniyle sadece ResNet34 mimarisinin evrişimli katmanları eğitilebilmiştir. Diğer sinir ağıları daha derin olduğundan az sayıda imge ile alt evrişimli katmanlar verimli bir şekilde eğitilememiştir. Ya sadece son katman eğitimi ile aynı başarı elde edilebilmiş ya da eğitim kümesi ezberlendiğinden test başarısı düşük kalmıştır.

En iyi sonuçlar ResNeXt50 sinir ağının evrişimli katmanları değiştirilmeden sadece son katman eğitilerek elde edilmiştir. Ancak farklı mimariler için elde edilen sonuçlar birbirine yakındır. Görece daha sığ olan ResNet34 sinir ağı da evrişimli katmanları eğitildiğinde diğerlerinin başarısına ulaşmaktadır. ResNeXt50 için karışıklık matrisi Şekil 3'te verilmiştir. Toplam üç hata bulunmaktadır: Sıfır numaralı doku, iki ve dört numaralı dokular ile, bir numaralı doku ise beş numaralı doku ile karıştırılmaktadır. Şekil 1 incelendiğinde bunların görece birbirine yakın dokular olduğu gözlemlenebilir.

64×64 imgeler için elde edilen sonuçlar Tablo II'de yer almaktadır. İmge boyutu artınca zemin dokusu sınıflandırma başarısı da artmaktadır. DenseNet121 sinir ağı sadece son katmanın eğitilmesi ile, ResNet34 sinir ağı ise evrişimli katmanların da eğitildiği durumda hatasız sınıflandırma yapmaktadır. Ayrıca genelde eğitim kayıp değerleri doğrulama kaybının üzerindedir. Bu da eğitim verisine aşırı uyum sorununun 32×32 imgelere göre görece daha az yaşandığına işaret

etmektedir.

İmge boyutu büyüyünce evrişimli katman filtrelerinin boyutları sabit olduğu için her bir filtre için daha fazla veri elde edilmiş olur. Bu sayede 64×64 imgeler kullanıldığında ResNeXt50 sinir ağının da alt katmanları eğitilebilmiştir ve sadece son katmanın eğitimine göre daha yüksek bir başarı elde edilmiştir. Ancak görece daha basit ResNet34 sinir ağının başarısına ulaşamamıştır. Bunun temel nedeninin veri azlığından dolayı alt evrişimli katmanların yeterince eğitilememesi olduğu düşünülmektedir. Ancak eğitim hızı bu katmanlar için daha da düşürülerek eğitim tekrarlandığında bir iyileşme gözlenmemiştir. Eğitim verisine daha fazla dönüşüm uygulanması ya da zemin dokularından daha fazla eğitim verisi toplanması alt katmanların eğitim başarısını artırabilir.

Deneyler sonunda daha önceden eğitilmiş derin sinir ağlarının yeni bir doku sınıflandırma uygulamasına adaptasyonunun mümkün olduğu, mimari derinliğin görece az olduğu ve yeterli büyüklükte imgeler kullanıldığı durumlarda tüm evrişimli katmanların eğitilebildiği gözlemlenmiştir. Bunun da başarıya olumlu etkisi olmuştur. Veri setinde çok farklı açılardan çekilmiş görüntüler olmasına rağmen yüksek başarı elde edilebilmiştir. Bu da en azından hazırlanan veri seti üzerinde zemin dokusu sınıflandırması için sıfırdan eğitim yapılması ya da Fisher vektörü hesaplanması benzeri adımların gerekmediğini göstermektedir.

IV. SONUÇ VE TARTIŞMA

Bu çalışmada önceden eğitilmiş derin sinir ağlarının zemin dokusu sınıflandırması için kısmen ya da tamamen eğitilmesi araştırılmıştır. Farklı sinir ağı mimari ve konfigürasyonları ile yapılan testler, önceden eğitilmiş sinir ağlarının başarıyla yeni bir doku sınıflandırması alanına aktarılabilirliğini göstermiştir. Doğrudan transfer öğrenmesi ile doku sınıflandırması, literatürde yer alan [4] ve [11] tarafından önerilen yöntemlerden kavramsal olarak daha basittir. Eğitim zamanları görece kısa ve az miktarda hafıza kullanımını gerektirmektedir.

Çalışmada sınırlı sayıda zemin dokusu kategorisi (yedi) ve görece az eğitim verisi (sınıf başına yetmiş adet) kullanılmıştır. Burada temel sınırlayıcı mevcut grafik kartı altyapısındaki eksikliklerdir. İleride veri setinin büyütülmesi ile çalışmanın özellikle daha büyük imgeler ile tekrarlanması planlanmaktadır. Bu da daha fazla mimari için evrişimli katmanların eğitilmesi sağlayacaktır. Özellikle mimari derinlik arttığında

TABLO I: 32×32 İMGELER İLE EĞİTİM SONUÇLARI

Mimari	Eğitilen Katmanlar	Eğitim Kayıp Değeri	Doğrulama Kayıp Değeri	Doğruluk (%)
ResNet34	Son Katman	0.0556	0.1549	95.7
	Tüm Katmanlar	0.0201	0.0931	96.4
ResNet50	Son Katman	0.0766	0.1923	96.4
	Tüm Katmanlar	-	-	-
DenseNet121	Son Katman	0.0389	0.0913	96.4
	Tüm Katmanlar	-	-	-
ResNeXt50	Son Katman	0.0972	0.1050	96.9
	Tüm Katmanlar	-	-	-

TABLO II: 64×64 İMGELER İLE EĞİTİM SONUÇLARI

Mimari	Eğitilen Katmanlar	Eğitim Kayıp Değeri	Doğrulama Kayıp Değeri	Doğruluk (%)
ResNet34	Son Katman	0.1480	0.0635	98.0
	Tüm Katmanlar	0.0118	0.0056	100.0
ResNet50	Son Katman	0.0705	0.0877	98.8
	Tüm Katmanlar	-	-	-
DenseNet121	Son Katman	0.0304	0.0226	100.0
	Tüm Katmanlar	-	-	-
ResNeXt50	Son Katman	0.0183	0.0542	98.4
	Tüm Katmanlar	0.0154	0.0206	99.2

veri setinin büyüklüğünün de artması gerekmektedir. Veri büyüklüğü yanında daha eğimli açılardan çekilmiş veya çevredeki nesnelerin zemini daha fazla örttüğü imgelerin veri setine katılması düşünülmektedir. Şu aşamada elde edilen sonuçlar daha zor veriler için de başarılı sınıflandırma için umut vaat etmektedir.

Çalışmanın geliştirilebileceği bir başka yön de eğitim metodolojisinin farklılaştırılmasıdır. Farklı eğitim algoritma ve yaklaşımlarının evrimsel katmanların eğitimi üzerine etkisi araştırılmalıdır. Ayrıca bir seferde eğitimde kullanılan örnek sayısı olan yığın boyutu da araştırılabilir.

Çalışmada yer alan sonuçlar evrimsel katmanların da eğitiminin yeterli veri olduğunda başarıyı arttırdığını göstermektedir. Eğitim sırasında bu başarı artışı görece kısa sürede gerçekleştiğinden, transfer öğrenmesi ile alt katman filtrelerin öğrenilmesinin en baştan eğitime göre daha kolay olduğu izlenimi vermektedir.

KAYNAKLAR

- [1] M. P. T. Ojala and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with Local Binary Patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.
- [2] M. Varma and A. Zisserman, "A statistical approach to texture classification from single images," *International Journal of Computer Vision*, vol. 62, no. 1-2, pp. 61–81, 2005.
- [3] M. Cimpoi, S. Maji, and A. Vedaldi, "Deep filter banks for texture recognition and segmentation," in *Conference on Computer Vision and Pattern Recognition*. IEEE, 2015, pp. 3828–3836.
- [4] V. Andrearczyk and P. F. Whelan, "Using filter banks in convolutional neural networks for texture classification," *Pattern Recognition Letters*, vol. 84, pp. 63–69, 2016.
- [5] X. Qi, C.-G. Li, G. Zhao, X. Hong, and M. Pietikäinen, "Dynamic texture and scene classification by transferring deep image features," *Neurocomputing*, vol. 171, pp. 1230–1241, 2016.
- [6] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.
- [7] Y. Bengio, "Deep learning of representations for unsupervised and transfer learning," in *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, 2012, pp. 17–36.
- [8] M. Varma and A. Zisserman, "Classifying images of materials: Achieving viewpoint and illumination independence," in *European Conference on Computer Vision*. Springer, 2002, pp. 255–271.
- [9] M. Unser, "Texture classification and segmentation using wavelet frames," *IEEE Transactions in Image Processing*, vol. 4, no. 11, pp. 1549–1560, 1995.
- [10] M. Varma and A. Zisserman, "Texture classification: Are filter banks necessary?" in *Conference on Computer Vision and Pattern Recognition*, vol. 2. IEEE, 2003, pp. II–691.
- [11] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014. [Online]. Available: <http://arxiv.org/abs/1409.1556>
- [12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [13] G. Huang, Z. Liu, K. Q. Weinberger, and L. van der Maaten, "Densely connected convolutional networks," in *Conference on Computer Vision and Pattern Recognition*, vol. 1, no. 2, 2017, p. 3.
- [14] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Conference on Computer Vision and Pattern Recognition*. IEEE, 2017, pp. 5987–5995.
- [15] A. Paszke, S. Gross, S. Chintala, and G. Chanan, "PyTorch: tensors and dynamic neural networks in python with strong gpu acceleration," <http://pytorch.org>, accessed: 2018-02-15.
- [16] J. Howard and R. Thomas, "fast.ai: making neural nets uncool again," <http://fast.ai>, accessed: 2018-02-15.
- [17] I. Loshchilov and F. Hutter, "SGDR: stochastic gradient descent with restarts," *CoRR*, vol. abs/1608.03983, 2016. [Online]. Available: <http://arxiv.org/abs/1608.03983>
- [18] L. N. Smith, "Cyclical learning rates for training neural networks," in *Winter Conference on Applications of Computer Vision*. IEEE, 2017, pp. 464–472.