Unsupervised Representation Learning With Deep Convolutional Generative Adversarial Networks

Alec Radford, Luke Metz and Soumith Chintala

International Conference on Learning Representations (ICLR), 2016

Problems Addressed

(a) Unsupervised learning using CNNs hasn't received much attention. Leveraging the huge amount of unlabeled images and videos to learn intermediate feature representations can be beneficial when used on supervised learning tasks like image classification.

(b) GANs have been unstable and difficult to train so far. The paper addresses this issue and improves the earlier architecture.

Proposed Solution

(a) Good image representations can be built by training GANs (Goodfellow et al., 2014), and later reusing parts of the generator and discriminator networks as feature extractors for supervised tasks.

(b) The deterministic spatial pooling functions in the networks (such as maxpooling) are replaced by strided convolutions. This allows the generator to learn its own spatial upsampling.

(c) Fully connected layers on top of convolutional features are removed. The highest convolutional features are directly connected to the input and output respectively of the generator and discriminator, as it improves model stability without much hurting the convergence speed.

(d) Batch normalization is adopted to stabilize learning by normalizing the input to each unit to have zero mean and unit variance. This helps deal with training problems that arise due to poor initialization and helps gradient flow in deeper models. It wasn't applied to generator output layer and discriminator input layer, however, as it lead to sample oscillation and model instability.

(e) Generator uses Tanh function in output layer and ReLU everywhere else. Discriminator uses LeakyReLU.

Claims

(a) No pre-processing was applied to training images besides scaling to the range of the tanh activation function [-1, 1].

(b) All models were trained with mini-batch stochastic gradient descent (SGD) with a mini-batch size of 128.

(c) All weights were initialized from a zero-centered Normal distribution with standard deviation 0.02. In the LeakyReLU, the slope of the leak was set to 0.2 in all models.

(d) Adam optimizer with tuned hyperparameters was used.

(e) The suggested learning rate of 0.001 was found to be too high, using 0.0002 instead.

(f) The momentum term B1 at the suggested value of 0.9 resulted in training oscillation and instability while reducing it to 0.5 helped stabilize training.

Results

(a) Samples from one epoch of training on the LSUN Bedrooms dataset are shown to demonstrate that this model is not producing high quality samples via simply overfitting/memorizing training examples. No data augmentation was applied to the images. To further increase likelihood of memorization a simple image de-duplication process was also performed.

(b) Training was done Imagenet-1k as a feature extractor and then the discriminator's features from all layers were fitted with linear L2-SVM classifier and run on CIFAR-10, which outperformed all K-Means based approaches with 82.8% accuracy.

(c) Interpolation in latent space resulted in semantic changes to the image generations (such as objects being added and removed), from which we can reason that the model has learned relevant and interesting representations.

(d) Only meaningful features are learnt, which can be reasoned from the observation that the network can be trained to "forget" specific features.

(e) Arithmetic operations were carried out on Z representation (of generators) of sets of exemplar samples for visual concepts. Experiments working on only single samples per concept were unstable, but averaging the Z vector for three examplars showed consistent and stable generations that semantically obeyed the arithmetic.

Bibliography

1. Bergstra, James and Bengio, Yoshua. Random search for hyper-parameter optimization. JMLR, 2012.
2. Coates, Adam and Ng, Andrew. Selecting receptive fields in deep networks. NIPS, 2011.
3. Coates, Adam and Ng, Andrew Y. Learning feature representations with k-means. In Neural Networks : Tricks of the Trade, pp. 561–580. Springer, 2012.
4. Deng, Jia, Dong, Wei, Socher, Richard, Li, Li-Jia, Li, Kai, and Fei-Fei, Li. Imagenet : A large-scale hierarchical image database. In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pp. 248–255. IEEE, 2009.
5. Denton, Emily, Chintala, Soumith, Szlam, Arthur, and Fergus, Rob. Deep generative image models using a laplacian pyramid of adversarial networks. arXiv preprint arXiv :1506.05751, 2015.
6. Dosovitskiy, Alexey, Springenberg, Jost Tobias, and Brox, Thomas. Learning to generate chairs with convolutional neural networks. arXiv preprint arXiv :1411.5928, 2014.
7. Dosovitskiy, Alexey, Fischer, Philipp, Springenberg, Jost Tobias, Riedmiller, Martin, and Brox, Thomas. Discriminative unsupervised feature learning with exemplar convolutional neural networks. In Pattern Analysis and Machine Intelligence, IEEE Transactions on, volume 99. IEEE, 2015.

8. Efros, Alexei, Leung, Thomas K, et al. Texture synthesis by non-parametric sampling. In Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on, volume 2, pp. 1033–1038. IEEE, 1999.
9. Freeman, William T, Jones, Thouis R, and Pasztor, Egon C. Example-based super-resolution. Computer Graphics and Applications, IEEE, 22(2) :56–65, 2002.
10. Goodfellow, Ian J, Warde-Farley, David, Mirza, Mehdi, Courville, Aaron, and Bengio, Yoshua. Maxout networks. arXiv preprint arXiv :1302.4389, 2013.
11. Goodfellow, Ian J., Pouget-Abadie, Jean, Mirza, Mehdi, Xu, Bing, Warde-Farley, David, Ozair, Sherjil, Courville, Aaron C., and Bengio, Yoshua. Generative adversarial nets. NIPS, 2014.
12. Gregor, Karol, Danihelka, Ivo, Graves, Alex, andWierstra, Daan. Draw : A recurrent neural network for image generation. arXiv preprint arXiv :1502.04623, 2015.
13. Hardt, Moritz, Recht, Benjamin, and Singer, Yoram. Train faster, generalize better : Stability of stochastic gradient descent. arXiv preprint arXiv :1509.01240, 2015.
14. Hauberg, Sren, Freifeld, Oren, Larsen, Anders Boesen Lindbo, Fisher III, John W., and Hansen, Lars Kair. Dreaming more data : Class-dependent distributions over diffeomorphisms for learned data augmentation. arXiv preprint arXiv :1510.02795, 2015.
15. Hays, James and Efros, Alexei A. Scene completion using millions of photographs. ACM Transactions on Graphics (TOG), 26(3) :4, 2007.
16. Ioffe, Sergey and Szegedy, Christian. Batch normalization : Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv :1502.03167, 2015.
17. Kingma, Diederik P and Ba, Jimmy Lei. Adam : A method for stochastic optimization. arXiv preprint arXiv :1412.6980, 2014.
18. Kingma, Diederik P and Welling, Max. Auto-encoding variational bayes. arXiv preprint arXiv :1312.6114, 2013.
19. Lee, Honglak, Grosse, Roger, Ranganath, Rajesh, and Ng, Andrew Y. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In Proceedings of the 26th Annual International Conference on Machine Learning, pp. 609–616. ACM, 2009.
20. Loosli, Ga¨elle, Canu, St´ephane, and Bottou, L´eon. Training invariant support vector machines using selective sampling. In Bottou, L´eon, Chapelle, Olivier, DeCoste, Dennis, and Weston, Jason (eds.), Large Scale Kernel Machines, pp. 301–320. MIT Press, Cambridge, MA., 2007. URL http ://leon.bottou.org/papers/loosli-canu-bottou-2006.
21. Maas, Andrew L, Hannun, Awni Y, and Ng, Andrew Y. Rectifier nonlinearities improve neural network acoustic models. In Proc. ICML, volume 30, 2013.
22. Mikolov, Tomas, Sutskever, Ilya, Chen, Kai, Corrado, Greg S, and Dean, Jeff. Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems, pp. 3111–3119, 2013.
23. Mordvintsev, Alexander, Olah, Christopher, and Tyka, Mike. Inceptionism : Going deeper into neural networks. http ://googleresearch.blogspot.com/2015/06/ inceptionism-going-deeper-into-neural.html. Accessed : 2015-06-17.
24. Nair, Vinod and Hinton, Geoffrey E. Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th International Conference on Machine Learning (ICML-10), pp. 807–814, 2010.

25. Netzer, Yuval, Wang, Tao, Coates, Adam, Bissacco, Alessandro, Wu, Bo, and Ng, Andrew Y. Reading digits in natural images with unsupervised feature learning. In NIPS workshop on deep learning and unsupervised feature learning, volume 2011, pp. 5. Granada, Spain, 2011.
26. Oquab, M., Bottou, L., Laptev, I., and Sivic, J. Learning and transferring mid-level image representations using convolutional neural networks. In CVPR, 2014.
27. Portilla, Javier and Simoncelli, Eero P. A parametric texture model based on joint statistics of complex wavelet coefficients. International Journal of Computer Vision, 40(1) :49–70, 2000.
28. Rasmus, Antti, Valpola, Harri, Honkala, Mikko, Berglund, Mathias, and Raiko, Tapani. Semisupervised learning with ladder network. arXiv preprint arXiv :1507.02672, 2015.
29. Sohl-Dickstein, Jascha, Weiss, Eric A, Maheswaranathan, Niru, and Ganguli, Surya. Deep unsupervised learning using nonequilibrium thermodynamics. arXiv preprint arXiv :1503.03585, 2015.
30. Springenberg, Jost Tobias, Dosovitskiy, Alexey, Brox, Thomas, and Riedmiller, Martin. Striving for simplicity : The all convolutional net. arXiv preprint arXiv :1412.6806, 2014.
31. Srivastava, Rupesh Kumar, Masci, Jonathan, Gomez, Faustino, and Schmidhuber, J¨urgen. Understanding locally competitive networks. arXiv preprint arXiv :1410.1165, 2014.
32. Theis, L., van den Oord, A., and Bethge, M. A note on the evaluation of generative models. arXiv :1511.01844, Nov 2015. URL http ://arxiv.org/abs/1511.01844.
33. Vincent, Pascal, Larochelle, Hugo, Lajoie, Isabelle, Bengio, Yoshua, and Manzagol, Pierre-Antoine. Stacked denoising autoencoders : Learning useful representations in a deep network with a local denoising criterion. The Journal of Machine Learning Research, 11 :3371–3408, 2010.
34. Xu, Bing, Wang, Naiyan, Chen, Tianqi, and Li, Mu. Empirical evaluation of rectified activations in convolutional network. arXiv preprint arXiv :1505.00853, 2015.
35. Yu, Fisher, Zhang, Yinda, Song, Shuran, Seff, Ari, and Xiao, Jianxiong. Construction of a large-scale image dataset using deep learning with humans in the loop. arXiv preprint arXiv :1506.03365, 2015.
36. Zeiler, Matthew D and Fergus, Rob. Visualizing and understanding convolutional networks. In Computer Vision–ECCV 2014, pp. 818–833. Springer, 2014.
37. Zhao, Junbo, Mathieu, Michael, Goroshin, Ross, and Lecun, Yann. Stacked what-where autoencoders. arXiv preprint arXiv :1506.02351, 2015.