[[1]](#footnote-1)

Generating building images with RIT features using DCGANs(November 2019)

Brian Landy, BSMS Computer Engineering, Rochester Institute of Technology

**Abstract—Deep Learning models can do many things that humans cannot. They can even extract features and reproduce fake data. Deep convolutional generative adversarial networks are an interesting part of the deep learning body of knowledge and this experiment using such models will attempt to demonstrate their use in creating believable generated images based on features of Rochester Institute of Technology buildings. Generative adversarial networks are a type of deep learning model which can be used to construct fake data for several different mediums. Deep convolutional generative adversarial networks are an extension of generative adversarial networks with the difference being that they use image convolutions to extract different types of features. Many images of the Rochester Institute of Technology campus were captured and processed in order to train a deep convolutional generative adversarial network. Several experiments with model changes were run and output was assessed visually at points in training. By varying some hyperparameters of the model from the baseline, it is shown much improvement can be made to output image quality.** **The task of generating truly believable output was much more difficult to deliver on.**

# INTRODUCTION

Performance in deep learning tasks continues to increase as more data is collected and used in training, as mentioned by Alom et al. [1]. In some cases, there aren’t enough samples to effectively train a model and so generative approaches are used to make more data. Goodfellow et al. [2] proposed a new type of generative network, the generative adversarial network (GAN), which is an unsupervised model where a generator network which tries to fool a discriminator network with fake samples. Since the first introduction of GANs, many variants have been proposed and successfully applied. GANs are computationally expensive and are even more so with the application of convolutional layers for feature extraction on images. The application of convolutional layers to GANs is introduced by Radford et al. [3] where the group proposes an architecture of GANs for better image generation. GANs have many parameters that can be altered to increase output quality, and better train the model, as demonstrated by some of the techniques presented by Salimans et al. [4]. A couple of these techniques were tried in the experiments of this project. A similar facade generation task was performed by Bachl et al. [5] and this paper has some similarities in methodology to the DCGAN attempts Bachl et al. [5] had made. The rest of this paper will cover topics in this order: the Background will explain DCGANs and the scope of the generation task, the Proposed Method section will discuss the choices made to tune the DCGAN model to generate better fake Rochester Institute of Technology (RIT) images, the results section will cover the best case outcomes and parameters and the conclusion will briefly restate findings and outcomes. Supplementary information about GANs and less noteworthy experiments will be in the appendix.

# Background

## DCGANs and Convolutional Feature Extraction

Radford et al. [3] proposed a DCGAN architecture than applies the image extraction properties of convolutional neural networks (CNN) to the generative ability of GANs. The discriminator network functioned similarly to a binary classifier convolutional neural network. To generate samples, transposed convolutional layers are employed by Radford et al. [3]. Transposed convolutions are opposite to convolutional layers in that they produce an output larger than the input and are meant for up sampling. Figure 1 shows the architecture chosen for the generator network in this experiment from Radford et al. [3].

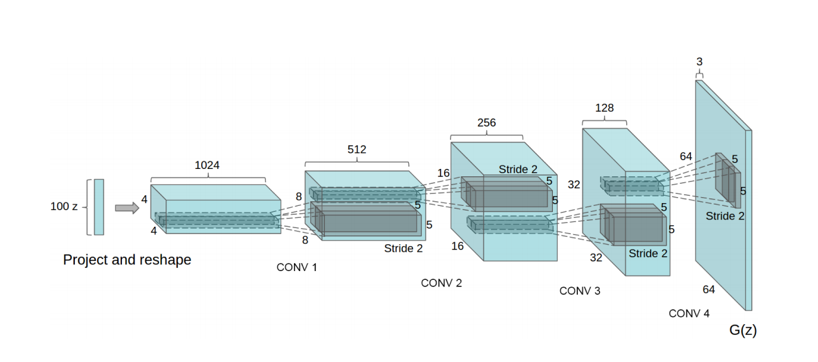


Fig. 1. Architecture for a Deep Convolutional Generator Network. In the figure, transposed convolutional layers and leaky ReLU activations transform random noise vectors into image features. Fig. 1 is from Radford et al. [3].

## Tools and Chosen Architecture

The chosen model to use as a baseline for this was the Pytorch DCGAN example from Radford et al. [6]. DCGAN was experimented with by Bachl et al. [5], but the goal in that paper was to generate new building images with datasets from many cities. The DCGAN model had trouble with the dataset and so some restrictions have been placed on the profile of building images allowed in the custom RIT dataset to negate these issues. They are listed in the datasets section of this paper. These images were processed in batches of 64 at a time.

## Datasets and Collection Methods

This project required the collection of images of RIT. Video was taken and the images were broken out frame by frame. This was done on a smartphone camera and sampled down to 64 pixel wide square images. The first captured set was of one side of the bioscience building and it is a minimally variational image set. By using this, the model should be able to easily overfit the set and produce one type of image. A larger, highly variational, set was also collected. The goal with this set was to capture front facing images, and to vary the type of buildings as the subject of the photos. This set has 29,297 images. LSUN churches is an image set of churches and was also used as a preliminary test with this DCGAN codebase.

## Modifications to DCGAN to Improve Output Believability

Many hyperparameters of the DCGAN model were modified. Some techniques used and assessed were: weakening the discriminator by lowering its learning rate, implementing soft labels, using flipped labels, swapping the activation functions to all leaky ReLU or changing the negative slope, varying the number of the generator filters, varying the adaptive moment estimation (Adam) optimizer regularization, lengthening the number of training epochs, and adding noisy, flipped, and rotated samples to provide a larger set to train on. Combinations of these were also tried More experiments were done and are mentioned in the appendix.

# Results

## Baseline Results on All Image Sets

Loss functions for D and G and prediction output of D for real and fake samples were plotted during training. Using an unmodified DCGAN [6] as a baseline, training on LSUN churches for 75 epochs produced the fake images seen in fig. 2.



Fig. 2. LSUN churches had fake images resembling churches but when looked at closely, it is still clear that they aren’t real. But from a distance and with a lower attention to detail, these are passable.

The low variation RIT set was run with the baseline for 560 epochs and generated images are shown in fig. 3.



Fig. 3. RIT low variation dataset. The output is very similar to the input and this is as expected due to the model overfitting the one building face.

The highly variable set was trained with, and at 223 epochs, the baseline DCGAN generated the images shown in fig. 4.



Fig. 4. RIT high variation dataset. The output is noisy, low quality, and clearly fake. The goal of the paper is to produce images that look better than this. G Filter depth: 64 G learning rate: 0.0002, D learning rate: 0.0002 Epoch: 223

## Troubleshooting Based on Loss Plots and D Output

Plotting loss proved to be an effective way to diagnose issues with the model early on ass suggested by Chintala et al[7]. If the D loss went to 0 way too early, then the discriminator became too good at classifying fakes and the generator loss kept increasing rapidly. The fake output samples looked terrible under these conditions. Over time G loss should keep increasing, but not radically, because it needs to do more and more work to fool a constantly trained discriminator until it satisfies its objective. The output of the discriminator was also plotted for real and fake samples to see if it was being fooled. But, since the dimensions of the input are so large, it is difficult for the generator to properly fool the discriminator, even when the model produces believable fake images. Due to this, the information was not as useful as hoped for. Ideally though, the output of D would end up moving away from 0 for fakes and away from 1 for real images as the classifier would start to be fooled. Both types of plots, loss and D output, are shown for reference in the appendix.

## Best Results of Modified Runs on RIT Datasets

The samples determined to be the best of each model were handpicked as every model had produced several bad samples. The best output produced came from training with a larger number of filters in the generator, training for more epochs and using a lower discriminator learning rate than the generator. Shown in fig. 5 is an example of best quality observed output.



Fig. 5. RIT High Variation Set Fake Samples Best Output. Filter depth: 128 G learning rate: 0.0002, D learning rate: 0.0001 Epoch: 680-700, and One-sided Noisy Labels [4]. Images were taken from a few different epochs

These images do a good job of conveying the small feature additions and disappearances that can happen when generating similar fake images. Some samples have trees and windows of a certain style and others do not. Again, these results had to be hand selected from all the generated images as each trained model still had good and bad outputs. But the more believable images generated in fig. 5 were better than the rest, visually speaking.

# Conclusion

After training these DCGANS and looking at results, it is apparent that there is a lot to the process of generating believable fakes. Aside from fig. 5, there were other models with decent fakes too, but they occurred less frequently and there were some models with very obviously fake output. Unfortunately, it seems that some of the more realistic fake images in most cases look very similar to real buildings from the input samples. A final observation is that the best-looking fakes looked like real samples but had small detail differences like missing features or extra features and were not radically different as hoped for. The output quality benefits greatly from modifying certain parameters over others, but when evaluating those changes, much time is spent pouring over large amounts of images that just barely scratch the surface of a convincing fake.

# References

[1] <https://arxiv.org/ftp/arxiv/papers/1803/1803.01164.pdf>

[2] <https://arxiv.org/pdf/1406.2661.pdf>

[3] <https://arxiv.org/pdf/1511.06434.pdf>

[4] <https://arxiv.org/pdf/1606.03498.pdf>

[5] <https://arxiv.org/pdf/1907.05280.pdf>

[6] <https://github.com/pytorch/examples/tree/master/dcgan>

[7] <https://github.com/soumith/ganhacks>

# Appendix

## Generative Adversarial Networks

Goodfellow et al. [2] presented the GAN which is made of two dueling networks. The networks used in the DCGAN are convolutional neural networks. A generator network takes a random input vector and passes a fake generated sample to a discriminator. The discriminator takes in real and generated samples and then determines whether a sample is real or fake. The training process is complex as it relies on the training of two networks, but if the discriminator detection confidence has been reduced to 50% for all samples (real and fake), then the generator has successfully fooled the discriminator. This section will outline important training equations for GANs as first described in Goodfellow et al. [2]. Discriminator networks are meant to maximize (1), which represents properly labeling a real sample *x* with a 1.

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

Generator networks are meant to minimize (2) which represents properly labeling a fake sample 0.

|  |  |  |
| --- | --- | --- |
|  |  | (2) |

These goals are meant to be met simultaneously by the networks. The general architecture for a GAN includes a real sample set, a random noise generator input, a generator network and a discriminator network. Shown in fig. 6 is the architecture for a standard GAN.



Fig. 6. Architecture for a GAN. In the figure, real samples, *x*,and fake samples, G(z), are passed into a discriminating network that will predict a label, D(G(z)) or D(x). The comparison between real and predicted labels are used to update discriminator and generator loss.

Goodfellow et al. [2] outlines the training process for GANs. Model updates occur in two steps. The discriminator network (D) is trained with *m* real and *m*fake samples. The weights of D are then adjusted using (3) which comes from Goodfellow et al. [2].

|  |  |  |
| --- | --- | --- |
|  |  | (3) |

The Generator network (G) is updated one time after updating D. This is shown in (4) which is also provided by Goodfellow et al. [2].

|  |  |  |
| --- | --- | --- |
|  |  | (4) |

The updates are performed with stochastic gradient descent and these are the gradients produced by the cost functions.

## Other Methods Used to Potentially Improve Output

**List all methods here**

## A Few of the Baseline Samples for Reference

To get an idea of the colors and quality of the regular training samples fig. 7 was included.



Fig. 7. RIT high variation dataset sample images.

## Plotted Loss and Discriminator Output

An example of the loss plot mentioned is shown in fig. 8.

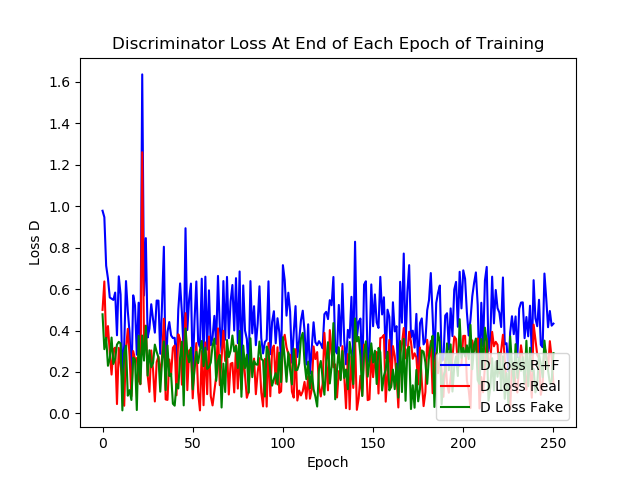


Fig. 8. D, G and the sum of D and G loss can be seen on the plot.

These helped a small amount and were at least interesting for trying to compare training behavior of different models. As for discriminator output, fig. 9. shows a typical discriminator output result during training.

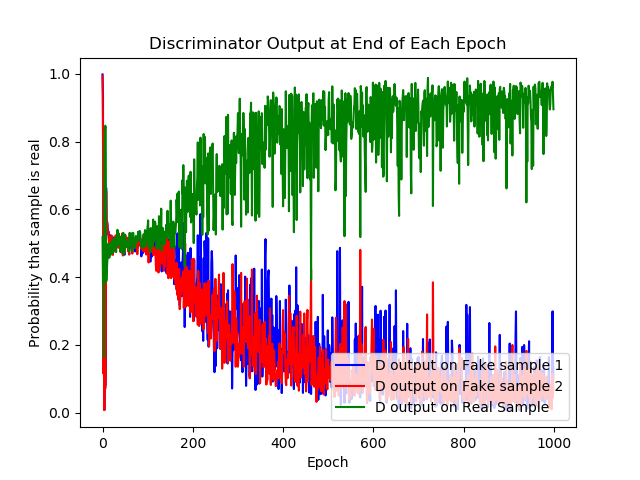


Fig. 9. Discriminator output during training process. Green is the real sample probability and red/blue is the fake.

## Training Environment and Process

The training took place on a machine with a 1070ti and cuda support. This helped speed up training dramatically and allowed for many different trials and combinations, with varied hyperparameters, to be run.

## Additional Options in Pytorch Script

**<<<I added stuff to the pytorch script to more easily try combinations of hyperparams together.**

## Techniques That Performed Much Worse than Expected

Since deep learning models benefit from more data, simple data augmentation was employed to created rotated, flipped and noisy samples. However, this dramatically worsened the generated images and wasn’t a good approach. The set was augmented by flipping images horizontally, rotating them 45 degrees and adding small amounts of noise. This had the opposite effect and output generated was some of the worst. The image set grew to around 100,000 samples.

References

*Basic format for books:*

J. K. Author, “Title of chapter in the book,” in *Title of His Published Book, x*th ed. City of Publisher, (only U.S. State), Country: Abbrev. of Publisher, year, ch. *x*, sec. *x*, pp. *xxx–xxx.*

*Examples:*

1. G. O. Young, “Synthetic structure of industrial plastics,” in *Plastics,* 2nd ed., vol. 3, J. Peters, Ed. New York, NY, USA: McGraw-Hill, 1964, pp. 15–64.
2. W.-K. Chen, *Linear Networks and Systems.* Belmont, CA, USA: Wadsworth, 1993, pp. 123–135.

*Basic format for periodicals:*

J. K. Author, “Name of paper,” *Abbrev. Title of Periodical*, vol. *x, no*. *x,* pp*. xxx-xxx,* Abbrev. Month, year, DOI. 10.1109.*XXX*.123456.

*Examples:*

1. J. U. Duncombe, “Infrared navigation—Part I: An assessment of feasibility,” *IEEE Trans. Electron Devices*, vol. ED-11, no. 1, pp. 34–39, Jan. 1959, 10.1109/TED.2016.2628402.
2. E. P. Wigner, “Theory of traveling-wave optical laser,”   
   *Phys. Rev*.,   
   vol. 134, pp. A635–A646, Dec. 1965.
3. E. H. Miller, “A note on reflector arrays,” *IEEE Trans. Antennas Propagat*., to be published.

*Basic format for reports:*

J. K. Author, “Title of report,” Abbrev. Name of Co., City of Co., Abbrev. State, Country, Rep. *xxx*, year.

*Examples:*

1. E. E. Reber, R. L. Michell, and C. J. Carter, “Oxygen absorption in the earth’s atmosphere,” Aerospace Corp., Los Angeles, CA, USA, Tech. Rep. TR-0200 (4230-46)-3, Nov. 1988.
2. J. H. Davis and J. R. Cogdell, “Calibration program for the 16-foot antenna,” Elect. Eng. Res. Lab., Univ. Texas, Austin, TX, USA, Tech. Memo. NGL-006-69-3, Nov. 15, 1987.

*Basic format for handbooks:*

*Name of Manual/Handbook, x* ed., Abbrev. Name of Co., City of Co., Abbrev. State, Country, year, pp. *xxx-xxx.*

*Examples:*

1. *Transmission Systems for Communications*, 3rd ed., Western Electric Co., Winston-Salem, NC, USA, 1985, pp. 44–60.
2. *Motorola Semiconductor Data Manual*, Motorola Semiconductor Products Inc., Phoenix, AZ, USA, 1989.

*Basic format for books (when available online):*

J. K. Author, “Title of chapter in the book,” in *Title of Published Book*, *x*th ed. City of Publisher, State, Country: Abbrev. of Publisher, year, ch. *x*, sec. *x*, pp. *xxx–xxx*. [Online]. Available: http://www.web.com

*Examples:*

1. G. O. Young, “Synthetic structure of industrial plastics,” in Plastics, vol. 3, Polymers of Hexadromicon, J. Peters, Ed., 2nd ed. New York, NY, USA: McGraw-Hill, 1964, pp. 15-64. [Online]. Available: http://www.bookref.com.
2. *The Founders’ Constitution*, Philip B. Kurland and Ralph Lerner, eds., Chicago, IL, USA: Univ. Chicago Press, 1987. [Online]. Available: http://press-pubs.uchicago.edu/founders/
3. The Terahertz Wave eBook. ZOmega Terahertz Corp., 2014. [Online]. Available: http://dl.z-thz.com/eBook/zomega\_ebook\_pdf\_1206\_sr.pdf. Accessed on: May 19, 2014.
4. Philip B. Kurland and Ralph Lerner, eds., *The Founders’ Constitution.* Chicago, IL, USA: Univ. of Chicago Press, 1987, Accessed on: Feb. 28, 2010, [Online] Available: http://press-pubs.uchicago.edu/founders/

*Basic format for journals (when available online):*

J. K. Author, “Name of paper,” *Abbrev. Title of Periodical*, vol. *x*, no. *x*, pp. *xxx-xxx*, Abbrev. Month, year. Accessed on: Month, Day, year, DOI: 10.1109.*XXX*.123456, [Online].

*Examples:*

1. J. S. Turner, “New directions in communications,” *IEEE J. Sel. Areas Commun*., vol. 13, no. 1, pp. 11-23, Jan. 1995.
2. W. P. Risk, G. S. Kino, and H. J. Shaw, “Fiber-optic frequency shifter using a surface acoustic wave incident at an oblique angle,” *Opt. Lett.*, vol. 11, no. 2, pp. 115–117, Feb. 1986.
3. P. Kopyt *et al., “*Electric properties of graphene-based conductive layers from DC up to terahertz range,” *IEEE THz Sci. Technol.,* to be published. DOI: 10.1109/TTHZ.2016.2544142.

*Basic format for papers presented at conferences (when available online):*

J.K. Author. (year, month). Title. presented at abbrev. conference title. [Type of Medium]. Available: site/path/file

*Example:*

1. PROCESS Corporation, Boston, MA, USA. Intranets: Internet technologies deployed behind the firewall for corporate productivity. Presented at INET96 Annual Meeting. [Online]. Available: http://home.process.com/Intranets/wp2.htp

*Basic format for reports and handbooks (when available online):*

J. K. Author. “Title of report,” Company. City, State, Country. Rep. no., (optional: vol./issue), Date. [Online] Available: site/path/file

*Examples:*

1. R. J. Hijmans and J. van Etten, “Raster: Geographic analysis and modeling with raster data,” R Package Version 2.0-12, Jan. 12, 2012. [Online]. Available: http://CRAN.R-project.org/package=raster
2. Teralyzer. Lytera UG, Kirchhain, Germany [Online]. Available: http://www.lytera.de/Terahertz\_THz\_Spectroscopy.php?id=home, Accessed on: Jun. 5, 2014

*Basic format for computer programs and electronic documents (when available online):*

Legislative body. Number of Congress, Session. (year, month day). *Number of bill or resolution*, *Title*. [Type of medium]. Available: site/path/file

***NOTE:*** ISO recommends that capitalization follow the accepted practice for the language or script in which the information is given.

*Example:*

1. U.S. House. 102nd Congress, 1st Session. (1991, Jan. 11). *H. Con. Res. 1, Sense of the Congress on Approval of Military Action*. [Online]. Available: LEXIS Library: GENFED File: BILLS

*Basic format for patents (when available online):*

Name of the invention, by inventor’s name. (year, month day). Patent Number[Type of medium]. Available: site/path/file

*Example:*

1. Musical toothbrush with mirror, by L.M.R. Brooks. (1992, May 19). Patent D 326 189

[Online]. Available: NEXIS Library: LEXPAT File: DES

*Basic format for conference proceedings (published):*

J. K. Author, “Title of paper,” in *Abbreviated Name of Conf.*, City of Conf., Abbrev. State (if given), Country, year, pp. *xxxxxx.*

*Example:*

1. D. B. Payne and J. R. Stern, “Wavelength-switched pas- sively coupled single-mode optical network,” in *Proc. IOOC-ECOC,* Boston, MA, USA,1985,   
   pp. 585–590.

*Example for papers presented at conferences (unpublished):*

1. D. Ebehard and E. Voges, “Digital single sideband detection for interferometric sensors,” presented at the *2nd Int. Conf. Optical Fiber Sensors,* Stuttgart, Germany, Jan. 2-5, 1984.

*Basic format for patents:*

J. K. Author, “Title of patent,” U.S. Patent *x xxx xxx*, Abbrev. Month, day, year.

*Example:*

1. G. Brandli and M. Dick, “Alternating current fed power supply,” U.S. Patent 4 084 217, Nov. 4, 1978.

*Basic format**for theses (M.S.) and dissertations (Ph.D.):*

a) J. K. Author, “Title of thesis,” M.S. thesis, Abbrev. Dept., Abbrev. Univ., City of Univ., Abbrev. State, year.

b) J. K. Author, “Title of dissertation,” Ph.D. dissertation, Abbrev. Dept., Abbrev. Univ., City of Univ., Abbrev. State, year.

*Examples:*

1. J. O. Williams, “Narrow-band analyzer,” Ph.D. dissertation, Dept. Elect. Eng., Harvard Univ., Cambridge, MA, USA, 1993.
2. N. Kawasaki, “Parametric study of thermal and chemical nonequilibrium nozzle flow,” M.S. thesis, Dept. Electron. Eng., Osaka Univ., Osaka, Japan, 1993.

*Basic format for the most common types of unpublished references:*

a) J. K. Author, private communication, Abbrev. Month, year.

b) J. K. Author, “Title of paper,” unpublished.

c) J. K. Author, “Title of paper,” to be published.

*Examples:*

1. A. Harrison, private communication, May 1995.
2. B. Smith, “An approach to graphs of linear forms,” unpublished.
3. A. Brahms, “Representation error for real numbers in binary computer arithmetic,” IEEE Computer Group Repository, Paper R-67-85.

*Basic formats for standards:*

a) *Title of Standard*, Standard number, date.

b) *Title of Standard*, Standard number, Corporate author, location, date.

*Examples:*

1. IEEE Criteria for Class IE Electric Systems, IEEE Standard 308, 1969.
2. Letter Symbols for Quantities, ANSI Standard Y10.5-1968.

*Article number in reference examples:*

1. R. Fardel, M. Nagel, F. Nuesch, T. Lippert, and A. Wokaun, “Fabrication of organic light emitting diode pixels by laser-assisted forward transfer,” *Appl. Phys. Lett.*, vol. 91, no. 6, Aug. 2007, Art. no. 061103.
2. J. Zhang and N. Tansu, “Optical gain and laser characteristics of InGaN quantum wells on ternary InGaN substrates,” *IEEE Photon. J.*, vol. 5, no. 2, Apr. 2013, Art. no. 2600111

*Example when using et al.:*

1. S. Azodolmolky *et al.*, Experimental demonstration of an impairment aware network planning and operation tool for transparent/translucent optical networks,” *J. Lightw. Technol.*, vol. 29, no. 4, pp. 439–448, Sep. 2011.

**First A. Author** (M’76–SM’81–F’87) and all authors may include biographies. Biographies are often not included in conference-related papers. This author became a Member (M) of IEEE in 1976, a Senior Member (SM) in 1981, and a Fellow (F) in 1987. The first paragraph may contain a place and/or date of birth (list place, then date). Next, the author’s educational background is listed. The degrees should be listed with type of degree in what field, which institution, city, state, and country, and year the degree was earned. The author’s major field of study should be lower-cased.

The second paragraph uses the pronoun of the person (he or she) and not the author’s last name. It lists military and work experience, including summer and fellowship jobs. Job titles are capitalized. The current job must have a location; previous positions may be listed without one. Information concerning previous publications may be included. Try not to list more than three books or published articles. The format for listing publishers of a book within the biography is: title of book (publisher name, year) similar to a reference. Current and previous research interests end the paragraph.

The third paragraph begins with the author’s title and last name (e.g., Dr. Smith, Prof. Jones, Mr. Kajor, Ms. Hunter). List any memberships in professional societies other than the IEEE. Finally, list any awards and work for IEEE committees and publications. If a photograph is provided, it should be of good quality, and professional-looking. Following are two examples of an author’s biography.

**Second B. Author** was born in Greenwich Village, New York, NY, USA in 1977. He received the B.S. and M.S. degrees in aerospace engineering from the University of Virginia, Charlottesville, in 2001 and the Ph.D. degree in mechanical engineering from Drexel University, Philadelphia, PA, in 2008.

From 2001 to 2004, he was a Research Assistant with the Princeton Plasma Physics Laboratory. Since 2009, he has been an Assistant Professor with the Mechanical Engineering Department, Texas A&M University, College Station. He is the author of three books, more than 150 articles, and more than 70 inventions. His research interests include high-pressure and high-density nonthermal plasma discharge processes and applications, microscale plasma discharges, discharges in liquids, spectroscopic diagnostics, plasma propulsion, and innovation plasma applications. He is an Associate Editor of the journal *Earth*, *Moon*, *Planets*, and holds two patents.

Dr. Author was a recipient of the International Association of Geomagnetism and Aeronomy Young Scientist Award for Excellence in 2008, and the IEEE Electromagnetic Compatibility Society Best Symposium Paper Award in 2011.

**Third C. Author, Jr.** (M’87) received the B.S. degree in mechanical engineering from National Chung Cheng University, Chiayi, Taiwan, in 2004 and the M.S. degree in mechanical engineering from National Tsing Hua University, Hsinchu, Taiwan, in 2006. He is currently pursuing the Ph.D. degree in mechanical engineering at Texas A&M University, College Station, TX, USA.

From 2008 to 2009, he was a Research Assistant with the Institute of Physics, Academia Sinica, Tapei, Taiwan. His research interest includes the development of surface processing and biological/medical treatment techniques using nonthermal atmospheric pressure plasmas, fundamental study of plasma sources, and fabrication of micro- or nanostructured surfaces.

Mr. Author’s awards and honors include the Frew Fellowship (Australian Academy of Science), the I. I. Rabi Prize (APS), the European Frequency and Time Forum Award, the Carl Zeiss Research Award, the William F. Meggers Award and the Adolph Lomb Medal (OSA).

1. [↑](#footnote-ref-1)