Improving Diffusion models with Discriminator Guidance

reproducibility study

Timo Kleger **Viktoria Sartor**

Abstract

In this paper we will reproduce the results of the Diffusion Guidance technique from Kim et al. [1] to improve the quality of the generated images from existing pretrained Diffusion models. Furthermore we try to even further improve the Discriminator Guidance for Diffusion models by applying the ensemble method in the discriminator guidance step. As a result we show that the results of the paper are valid and we could also reconstruct it by our own even though we are slightly worse that the results from the paper. The ensemble method that we suggested did not improve the prediction as we expected. All our code that we used for generating, training and the experiments is available under `https://github.com/ChlaegerIO/KTH_DeepLearning-diffusion`.

1 Introduction

This study is about the famous Diffusion models [2] first introduced in 2015 that generate astonishing images from noise. In our highly digital world it is important to be able to generate images, because sometimes it is economically too expensive to produce image data to train our Artificial Intelligent systems or the generated images could advance or extend a dataset. Another important point that we address in this report is that the results should be reproducible. In this report we dive into the the challenge of reproducibility in Science and especially in Machine Learning as J. Pineau et. al. [3] stated in a recent paper. We will present our reproduciblity results of the paper and discuss a potential own improvement.

In recent years the diffusion models were improved drastically. In this report we focus on Discriminator Guided Diffusion models as stated in [1], which can be used as an extension on already fully trained Diffusion models without requiring to retrain or fine tune the Diffusion model. Our goal is to replicate the results from Kim et al. [1] on the CIFAR10 data set as we did not have the computational power to train the discriminator on larger data sets such as ImageNet 256x256. Our task consists of training a discriminator, generating images and then comparing the results, thus FID score, to the original paper [1]. Furthermore we show generated images with or without Discriminator Guidance.

In addition to the replication we experiment with Ensemble method to further improve the discriminator guidance step and thus to improve the quality of the image generation. We train several Ensemble members and then average over all predicted discrimination.

In our studies we show that the FID score improves from 2.09 to 1.88 with the pretrained model of the original paper. Our own trained Discriminator achieves an FID of 2.08 and the Ensemble Discriminator has a worse score of 17.93.

Project report for DD2412 or FDD3412

2 Related Work

2.1 Diffusion models

Diffusion models in computer vision, particularly Denoising Diffusion models, have shown promising results in generative modeling. Based on a forward diffusion stage and a reverse diffusion stage, these models have been applied in various frameworks [4], including Denoising Diffusion Probabilistic Models [5]. Despite their computational complexity, recent research has focused on making these models more efficient, particularly emphasizing design strategies that improve their computational efficiency [6].

Denoising Diffusion Probabilistic Models Ho et al. [5] introduced denoising diffusion probabilistic models, a class of latent variable models, and demonstrated their effectiveness in image synthesis. They achieved high-quality results on the CIFAR10 dataset and LSUN, with an Inception score of 9.46 and a state-of-the-art FID score of 3.17.

2.2 Discriminator Guidance

Kim et al. [1] proposes a method called Discriminator Guidance that improves sample generation in pretrained Diffusion models by incorporating a discriminator for realistic sample supervision. This approach achieves state-of-the-art results on ImageNet 256x256 without requiring joint training of score and discriminator networks. To understand our study we want to introduce the forward and especially the backward path of the Stochastic Differential Equation (SDE). In the forward path we add step by step noise to the image. In the backward path we learn to revert this process. In most cases the noise that was added to the image in the forward path is learned and then subtracted from the image, $\mathbf{s}_\theta$. In or case we further guide the path with a discriminator that distinguishes real from fake images, $\mathbf{d}_\phi$. This is visualized in figure 1. The continuous mathematical formulation of the forward SDE is

$$d\mathbf{x}_t = \mathbf{f}(\mathbf{x}_t, t)dt + g(t)d\mathbf{w} \tag{1}$$

and of the backward SDE it is

$$d\mathbf{x}_t = \left[\mathbf{f}(\mathbf{x}_t, t) - g^2(t)(\mathbf{s}_\theta + w_t^{(DG)}\mathbf{d}_\phi)(\mathbf{x}_t, t)\right] d\bar{t} + g(t)d\bar{\mathbf{w}}_t. \tag{2}$$

Where $\mathbf{f}(\mathbf{x}_t, t)$ and $g(t)$ are the drift and the volatility coefficients, dt and dw are the infinitesimal small time step and Brownian motion, thus where noise is added, and $\mathbf{s}_\theta$, $w_t^{(DG)}$ and $\mathbf{d}_\phi$ are the reverse diffusion score, weight of the Discriminator Guidance and Discriminator Guidance value given the parameters θ and ϕ.

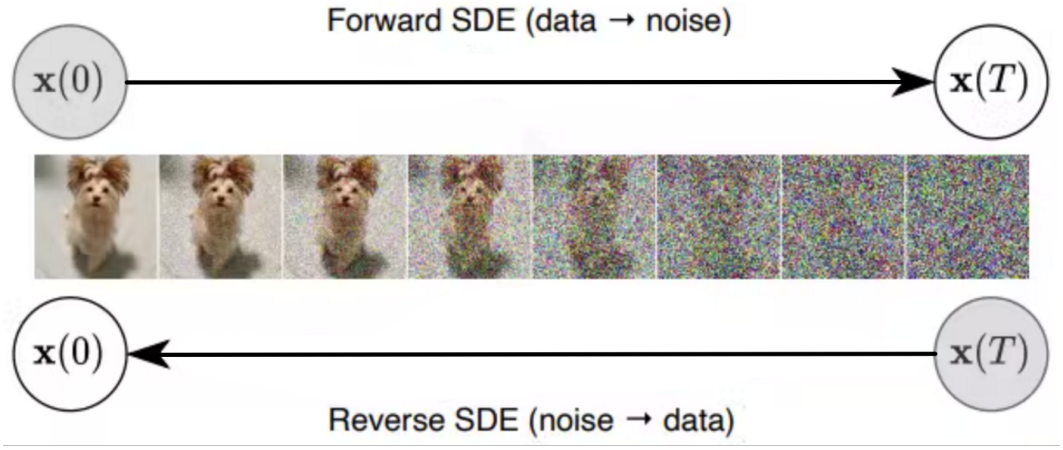

Figure 1: Forward and backward pass of diffusion models. The image is adapted from Song et al. [7].

Building upon this concept, Ho et al. [5] introduces Classifier-Free Diffusion Guidance. They demonstrate that guidance can be performed without a classifier by jointly training conditional and unconditional diffusion models.

Lim et al. [8] extend the application of score-based diffusion models to function spaces by introducing Denoising Diffusion Operators for training in these spaces. They demonstrate the applicability of these models in scientific computing and 3D geometric data analysis.

2.3 Discriminator Ensemble

Xu et al. [9] developed FairGAN, a fairness-aware generative adversarial network that generates discrimination-free data while preserving data utility. The model uses two discriminators to create fairness. Another model called PATE-GAN was developed by Yoon et al. [10]. It uses several discriminators, each trained discriminator ensure the possibility to generate synthetic data with differential privacy guarantees. It uses the Private Aggregation of Teacher Ensembles method.

3 Method

We focus on reproducing the results from the EDM-G++ model from Kim et al. [1]. This model uses the pretrained EDM Diffusion model. In this chapter we describe which data we use 3.1, the three steps how we proceed to get the reproduction results 3.2 - 3.4 and how we build our Ensemble 3.5 EDM-G++ model.

3.1 Data

The paper was tested on CIFAR10, CelebA, FFHQ and ImageNet256. Based on our computation power we decide to use CIFAR10 dataset [11] to reproduce the Discriminator Guidance method and to experiment with ensemble method. We also decided to test the unconditional case of the CIFAR10 data set. This is a harder task than the conditional one as we have less information about the image. The CIFAR10 data consists of 50'000 images in a resolution of (32,32,3) pixels. We have a dataloader which takes 50'000 fake images and 50'000 CIFAR10 images, then randomly shuffle the samples before we split the data set into train set, validation set and test set in the proportion of 80%, 10% and 10%.

3.2 Generate fake sample

For the training of the Discriminator we need generated fake samples. As we have the pretrained Diffusion model $\mathbf{s}_\theta$ we diffuse noisy randomly sampled images, $x_T \sim \mathcal{N}(0, I)$, to fake images via the diffusion path of equation 2 by setting the DG term to zero, $w_t^{(DG)} = 0$.

3.3 Discriminator Training

In the second step we train or in our case fine tune the Discriminator, which distinguishes real from fake images. The whole network consists of a pretrained and fixed classifier, which is a UNet architecture. This classifier takes the CIFAR10 images as an input, (B,32,32,3) and outputs a feature representation of it, (B,8,8,512). The second stage is an own or also a pretrained discriminator, to save computational power we also took a pretrained UNet model. This discriminator is now fine tuned to take the features and output a single number, (B, 1). We use a batch size B of 64 instead of the 128 used in the paper [1] as our hardware only supported 64. During training we closely measured accuracy and loss, both for the training data and for the validation data.

3.4 Generate Discriminator Guided samples and evaluation

For the evaluation FID score we need samples with the trained Discriminator. To generate these samples with Discriminator Guidance we use the parameters of table 1. The only difference is that with Discriminator Guidance we use a first order weight of 2 and boosting as it was done in the paper [1].

Table 1: Comparison of fake sample generation to sample generation with Discriminator Guidance

Parameter	**Fake Samples**	**DG Sample Generation**
Number of Diffusion steps	35	35
Minimum distance ($\times 10^{-5}$)	10	10
Maximum distance	$1 - 10 \times 10^{-5}$	$1 - 10 \times 10^{-5}$
Image size	32	32
DG weight 1st order	0	0
DG weight 2nd order	0	2.0
Time minimum	0.01	0.01
Time maximum	1.0	1.0
Boosting	False	True
Batch size	64	64
Number of samples	50,000	50,000

To evaluate our results we used FID [12], precision and recall, similar to the original paper. The results are in section 4

3.5 Ensembled Discriminator Guidance

To test our idea to use the ensemble method we trained six discriminators. The first ensemble is equal to the one in the paper and can thus be used for the reproducibility evaluation. As a first intuition if our idea works we just used different hyperparameters as seen in table 2. However is is recommended to have a bigger variance in the model selection to get better results and thus also use different architectures as ensembles and not only different hyperparameters. In table 2 we list all Discriminator Guidance hyperparameters that are relevant and different among the ensembles.

Table 2: Ensembled Discriminator Guidance hyperparameters used in training

Discriminator	Paper	D1	D2	D3	D4	D5
Number of Epochs	60	40	40	40	40	200
Learning Rate	0.0003	0.001	0.0001	0.0001	0.00005	0.00001
Weight Decay	1×10^{-7}	1×10^{-7}	0	1×10^{-3}	1×10^{-9}	1×10^{-11}
Min Difference Time	1×10^{-5}	1×10^{-5}	0.01	1×10^{-3}	1×10^{-3}	1×10^{-5}

To sample images with ensembled Discriminator Guidance we averaged the output of all six discriminator values.

———- add formula justification??? —-

4 Experiments and results

In the this chapter we explain the experiments and we state our results that we achieved in table 3.

Table 3: Evaluation results of reproduction and of ensemble addition

	EDM	EDM-G++	Our EDM-G++	Our ensemble EDM-G++
FID	2.094	1.880	2.084	17.933
Recall	-	0.761	0.733	0.736
Precission	-	0.500	0.525	0.446

4.1 Reproduction

For the FID we get 2.34....., which is quite a lot higher than the score in the paper

4.2 Ensemble method for discriminator guidance

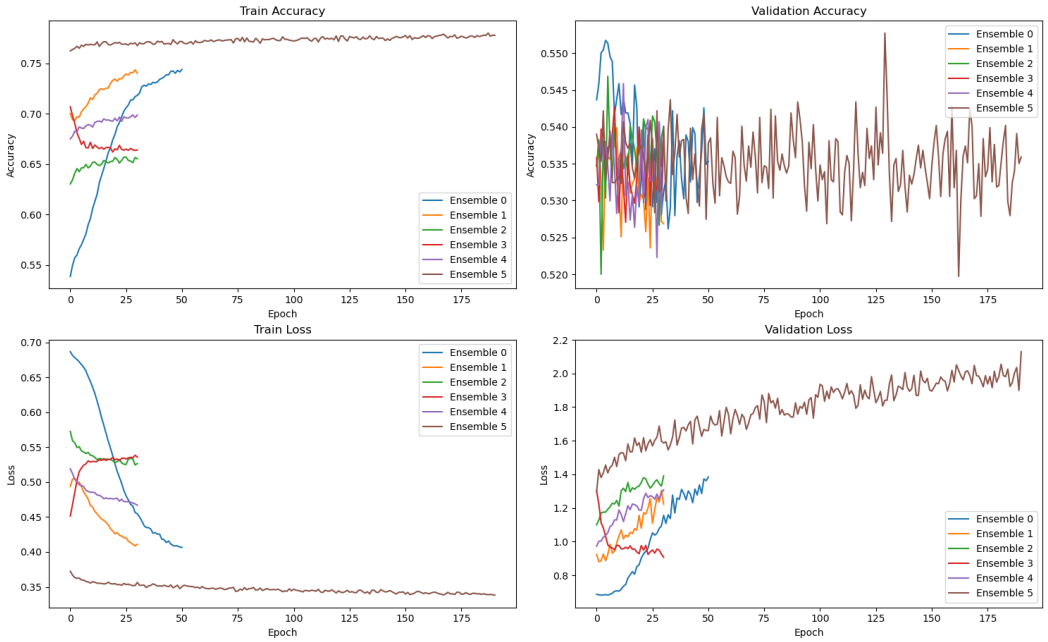

Figure 2: Discriminator Training

Empirically, ensembles tend to yield better results when there is a significant diversity among the models.

Frechet Inception Distance (FID) for unconditional case, daher schlechter??? Ist zwar 1.77 im paper tabelle 3!!!!

5 Challenges

(zusammen)
Fist of all, Diffusion models was a new field for all of us before this course. Therefore we successfully managed this interesting challenge to understand this field. Then a even bigger challenge was that one of our team mates got sick and gave up after the first xxxxxxxxxxx weeks. So we finished the project with only two persons.

6 Conclusion

(zusammen)

- Wieso haben sie kein validation set genommen

7 Ethical consideration, societal impact, alignment with UN SDG targets

(Vik: UN SDG targets, Timo: societal impacts) The sustainability development goal number 17 is a good fit, since we come from two different countries, working together on this project in Sweden[1].

8 Self Assessment

(zusammen)

[1] `https://sdgs.un.org/goals/goal17`

9 Acronym

Adam	Adaptive moment estimation
AI	Artificial Intelligence
DG	Discriminator Guidance
EDM	Equivariant Diffusion Model
EDM-G++	Equivariant Diffusion Model with DG
FID	Frechet Inception Distance
SDE	Stochastic Differential Equation
UN SDG	United Nations Sustainable Development Goals

References

[1] Dongjun Kim, Yeongmin Kim, Se Jung Kwon, et al. Refining generative process with discriminator guidance in score-based diffusion models, 2023. URL https://icml.cc/virtual/2023/oral/25468. ICML 2023.

[2] Sohl-Dickstein, Jascha, Weiss Eric, et al. Deep unsupervised learning using nonequilibrium thermodynamics. *PMLR. 37: 2256–2265*, 2015.

[3] J. Pineau et al. Improving reproducibility in machine learning research (a report from the neurips 2019 reproducibility program). In *arXiv:2003.12206*, 2020.

[4] Florinel-Alin Croitoru, Vlad Hondru, Radu Tudor Ionescu, and Mubarak Shah. Diffusion models in vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45:10850–10869, 2022. URL https://api.semanticscholar.org/CorpusID:252199918.

[5] Jonathan Ho, Ajay Jain, and P. Abbeel. Denoising diffusion probabilistic models. *ArXiv*, abs/2006.11239, 2020. URL https://api.semanticscholar.org/CorpusID:219955663.

[6] Anwaar Ulhaq, Naveed Akhtar, and Ganna Pogrebna. Efficient diffusion models for vision: A survey. *ArXiv*, abs/2210.09292, 2022. URL https://api.semanticscholar.org/CorpusID:252918532.

[7] Yang Song, Jascha Sohl-Dickstein, et al. Score-based generative modeling through stochastic differential equations. *ICLR (oral)*, 2021.

[8] Jae Hyun Lim, Nikola B. Kovachki, Ricardo Baptista, Christopher Beckham, Kamyar Azizzadenesheli, Jean Kossaifi, Vikram S. Voleti, Jiaming Song, Karsten Kreis, Jan Kautz, Christopher Joseph Pal, Arash Vahdat, and Anima Anandkumar. Score-based diffusion models in function space. *ArXiv*, abs/2302.07400, 2023. URL https://api.semanticscholar.org/CorpusID:256868496.

[9] Depeng Xu, Shuhan Yuan, Lu Zhang, and Xintao Wu. Fairgan: Fairness-aware generative adversarial networks. *2018 IEEE International Conference on Big Data (Big Data)*, pages 570–575, 2018. URL https://api.semanticscholar.org/CorpusID:44106659.

[10] Jinsung Yoon, James Jordon, and Mihaela van der Schaar. PATE-GAN: Generating synthetic data with differential privacy guarantees. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=S1zk9iRqF7.

[11] Alex Krizhevsky, Vinod Nair, and Geoffrey Hinton. Cifar-10 (canadian institute for advanced research). URL http://www.cs.toronto.edu/~kriz/cifar.html.

[12] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium, 2018.

10 Submission of report to DD2424

We require an electronic submission to the Canvas webpage. Please read the instructions below carefully and follow them faithfully.

10.1 Style

Papers to be submitted to NeurIPS 2018 must be prepared according to the instructions presented here. Papers may only be up to eight pages long, including figures. Additional pages *containing only acknowledgments and/or cited references* are allowed. Papers that exceed eight pages of content (ignoring references) will not be reviewed, or in any other way considered for presentation at the conference. The margins in 2018 are the same as since 2007, which allow for $\sim 15\%$ more words in the paper compared to earlier years. Authors are required to use the NeurIPS LaTeX style files obtainable at the NeurIPS website as indicated below. Please make sure you use the current files and not previous versions. Tweaking the style files may be grounds for rejection.

10.2 Retrieval of style files

The style files for NeurIPS and other conference information are available on the World Wide Web at

`http://www.neurips.cc/`

The file `neurips_2018.pdf` contains these instructions and illustrates the various formatting requirements your NeurIPS paper must satisfy. The only supported style file for NeurIPS 2018 is `neurips_2018.sty`, rewritten for LaTeX 2_ε. **Previous style files for LaTeX 2.09, Microsoft Word, and RTF are no longer supported!** The LaTeX style file contains three optional arguments: `final`, which creates a camera-ready copy, `preprint`, which creates a preprint for submission to, e.g., arXiv, and `nonatbib`, which will not load the `natbib` package for you in case of package clash.

New preprint option for 2018 If you wish to post a preprint of your work online, e.g., on arXiv, using the NeurIPS style, please use the `preprint` option. This will create a nonanonymized version of your work with the text "Preprint. Work in progress." in the footer. This version may be distributed as you see fit. Please **do not** use the `final` option, which should **only** be used for papers accepted to NeurIPS. At submission time, please omit the `final` and `preprint` options. This will anonymize your submission and add line numbers to aid review. Please do *not* refer to these line numbers in your paper as they will be removed during generation of camera-ready copies. The file `neurips_2018.tex` may be used as a "shell" for writing your paper. All you have to do is replace the author, title, abstract, and text of the paper with your own. The formatting instructions contained in these style files are summarized in Sections 11, 12, and 13 below.

11 General formatting instructions

The text must be confined within a rectangle 5.5 inches (33 picas) wide and 9 inches (54 picas) long. The left margin is 1.5 inch (9 picas). Use 10 point type with a vertical spacing (leading) of 11 points. Times New Roman is the preferred typeface throughout, and will be selected for you by default. Paragraphs are separated by ½ line space (5.5 points), with no indentation. The paper title should be 17 point, initial caps/lower case, bold, centered between two horizontal rules. The top rule should be 4 points thick and the bottom rule should be 1 point thick. Allow ¼ inch space above and below the title to rules. All pages should start at 1 inch (6 picas) from the top of the page. For the final version, authors' names are set in boldface, and each name is centered above the corresponding address. The lead author's name is to be listed first (left-most), and the co-authors' names (if different address) are set to follow. If there is only one co-author, list both author and co-author side by side. Please pay special attention to the instructions in Section 13 regarding figures, tables, acknowledgments, and references.

12 Headings: first level

All headings should be lower case (except for first word and proper nouns), flush left, and bold. First-level headings should be in 12-point type.

12.1 Headings: second level

Second-level headings should be in 10-point type.

12.1.1 Headings: third level

Third-level headings should be in 10-point type.

Paragraphs There is also a `\paragraph` command available, which sets the heading in bold, flush left, and inline with the text, with the heading followed by 1 em of space.

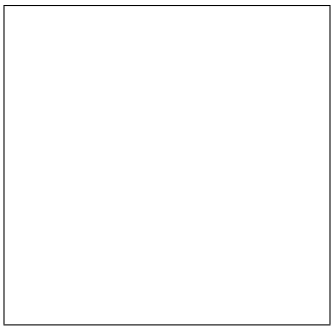

Figure 3: Sample figure caption.

13 Citations, figures, tables, references

These instructions apply to everyone.

13.1 Citations within the text

The `natbib` package will be loaded for you by default. Citations may be author/year or numeric, as long as you maintain internal consistency. As to the format of the references themselves, any style is acceptable as long as it is used consistently. The documentation for `natbib` may be found at

`http://mirrors.ctan.org/macros/latex/contrib/natbib/natnotes.pdf`

Of note is the command `\citet`, which produces citations appropriate for use in inline text. For example,

```
\citet{hasselmo} investigated\dots
```

produces

> Hasselmo, et al. (1995) investigated…

If you wish to load the `natbib` package with options, you may add the following before loading the `neurips_2018` package:

```
\PassOptionsToPackage{options}{natbib}
```

If `natbib` clashes with another package you load, you can add the optional argument `nonatbib` when loading the style file:

```
\usepackage[nonatbib]{neurips_2018}
```

As submission is double blind, refer to your own published work in the third person. That is, use "In the previous work of Jones et al. [4]," not "In our previous work [4]." If you cite your other papers that are not widely available (e.g., a journal paper under review), use anonymous author names in the citation, e.g., an author of the form "A. Anonymous."

13.2 Footnotes

Footnotes should be used sparingly. If you do require a footnote, indicate footnotes with a number[2] in the text. Place the footnotes at the bottom of the page on which they appear. Precede the footnote with a horizontal rule of 2 inches (12 picas). Note that footnotes are properly typeset *after* punctuation marks.[3]

13.3 Figures

All artwork must be neat, clean, and legible. Lines should be dark enough for purposes of reproduction. The figure number and caption always appear after the figure. Place one line space before the figure caption and one line space after the figure. The figure caption should be lower case (except for first word and proper nouns); figures are numbered consecutively. You may use color figures. However, it is best for the figure captions and the paper body to be legible if the paper is printed in either black/white or in color.

[2] Sample of the first footnote.

[3] As in this example.

Table 4: Sample table title

	Part	
Name	Description	Size (μm)
Dendrite	Input terminal	$\sim$100
Axon	Output terminal	$\sim$10
Soma	Cell body	up to 10^6

13.4 Tables

All tables must be centered, neat, clean and legible. The table number and title always appear before the table. See Table 4. Place one line space before the table title, one line space after the table title, and one line space after the table. The table title must be lower case (except for first word and proper nouns); tables are numbered consecutively. Note that publication-quality tables *do not contain vertical rules.* We strongly suggest the use of the `booktabs` package, which allows for typesetting high-quality, professional tables:

`https://www.ctan.org/pkg/booktabs`

This package was used to typeset Table 4.

14 Final instructions

Do not change any aspects of the formatting parameters in the style files. In particular, do not modify the width or length of the rectangle the text should fit into, and do not change font sizes (except perhaps in the **References** section; see below). Please note that pages should be numbered.

15 Preparing PDF files

Please prepare submission files with paper size "US Letter," and not, for example, "A4." Fonts were the main cause of problems in the past years. Your PDF file must only contain Type 1 or Embedded TrueType fonts. Here are a few instructions to achieve this.

- You should directly generate PDF files using `pdflatex`.
- You can check which fonts a PDF files uses. In Acrobat Reader, select the menu Files>Document Properties>Fonts and select Show All Fonts. You can also use the program `pdffonts` which comes with `xpdf` and is available out-of-the-box on most Linux machines.
- The IEEE has recommendations for generating PDF files whose fonts are also acceptable for NeurIPS. Please see `http://www.emfield.org/icuwb2010/downloads/IEEE-PDF-SpecV32.pdf`
- `xfig` "patterned" shapes are implemented with bitmap fonts. Use "solid" shapes instead.
- The `\bbold` package almost always uses bitmap fonts. You should use the equivalent AMS Fonts:

  ```
  \usepackage{amsfonts}
  ```

  followed by, e.g., `\mathbb{R}`, `\mathbb{N}`, or `\mathbb{C}` for $\mathbb{R}$, $\mathbb{N}$ or $\mathbb{C}$. You can also use the following workaround for reals, natural and complex:

  ```
  \newcommand{\RR}{I\!\!R} %real numbers
  \newcommand{\Nat}{I\!\!N} %natural numbers
  \newcommand{\CC}{I\!\!\!\!C} %complex numbers
  ```

  Note that `amsfonts` is automatically loaded by the `amssymb` package.

If your file contains type 3 fonts or non embedded TrueType fonts, we will ask you to fix it.

15.1 Margins in LaTeX

Most of the margin problems come from figures positioned by hand using `\special` or other commands. We suggest using the command `\includegraphics` from the `graphicx` package. Always specify the figure width as a multiple of the line width as in the example below:

```
\usepackage[pdftex]{graphicx} ...
\includegraphics[width=0.8\linewidth]{myfile.pdf}
```

See Section 4.4 in the graphics bundle documentation (`http://mirrors.ctan.org/macros/latex/required/graphics/grfguide.pdf`) A number of width problems arise when LaTeX cannot properly hyphenate a line. Please give LaTeX hyphenation hints using the `\-` command when necessary.

Acknowledgments

Use unnumbered third level headings for the acknowledgments. All acknowledgments go at the end of the paper. Do not include acknowledgments in the anonymized submission, only in the final paper.

References

References follow the acknowledgments. Use unnumbered first-level heading for the references. Any choice of citation style is acceptable as long as you are consistent. It is permissible to reduce the font size to `small` (9 point) when listing the references. **Remember that you can use more than eight pages as long as the additional pages contain *only* cited references.** [1] Alexander, J.A. & Mozer, M.C. (1995) Template-based

algorithms for connectionist rule extraction. In G. Tesauro, D.S. Touretzky and T.K. Leen (eds.), *Advances in Neural Information Processing Systems 7*, pp. 609–616. Cambridge, MA: MIT Press. [2] Bower, J.M. & Beeman, D. (1995) *The Book of GENESIS: Exploring Realistic Neural Models with the GEneral NEural SImulation System.* New York: TELOS/Springer–Verlag. [3] Hasselmo, M.E., Schnell, E. & Barkai, E. (1995) Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3. *Journal of Neuroscience* **15**(7):5249-5262.