staatliche
kunsthalle
baden-baden

MARTIN WÜHLER

SYNTHETIC MEDIA OR AN OBJECT THAT NEVER EXISTED

45cbm |Staatliche Kunsthalle Baden-Baden | 18.01. – 29.03.2020

45 cbm
FREUNDE
DER STAATLICHEN
KUNSTHALLE
BADEN-BADEN E.V.

CONTENT AWARE FILL

ASTROTURFING GAN

DARK PR FACE SWAP

IMAGE PHYLOGENY

DIGITAL

FORENSICS

1 introduction

The promise of deep learning is to discover rich, hierarchical models [2] that represent probability distributions over the kinds of data encountered in artificial intelligence applications, such as natural images, audio waveforms containing speech, and symbols in natural language corpora. So far, the most striking successes in deep learning have involved discriminative models, usually those that map a high-dimensional, rich sensory input to a class label [14, 22]. These striking successes have primarily been based on the backpropagation and dropout algorithms, using piecewise linear units [19, 9, 10] which have a particularly well-behaved gradient . Deep generative models have had less of an impact, due to the difficulty of approximating many intractable probabilistic computations that arise in maximum likelihood estimation and related strategies, and due to difficulty of leveraging the benefits of piecewise linear units in the generative context. We propose a new generative model estimation procedure that sidesteps these difficulties.

In the proposed adversarial nets framework, the generative model is pitted against an adversary: a discriminative model that learns to determine whether a sample is from the model distribution or the data distribution. The generative model can be thought of as analogous to a team of counterfeiters, trying to produce fake currency and use it without detection, while the discriminative model is analogous to the police, trying to detect the counterfeit currency. Competition in this game drives both teams to improve their methods until the counterfeits are indistiguishable from the genuine articles.

Jean Pouget-Abadie is visiting Universite de Montre al from Ecole Polytechnique.
†Sherjil Ozair is visiting Universite´ de Montre´ al from Indian Institute of Technology Delhi
‡Yoshua Bengio is a CIFAR Senior Fellow.
1 All code and hyperparameters available at http://www.github.com/goodfeli/adversarial

This framework can yield specific training algorithms for many kinds of model and optimization algorithm. In this article, we explore the special case when the generative model generates samples by passing random noise through a multilayer perceptron, and the discriminative model is also a multilayer perceptron. We refer to this special case as adversarial nets. In this case, we can train both models using only the highly successful backpropagation and dropout algorithms [17] and sample from the generative model using only forward propagation. No approximate inference or Markov chains are necessary.

2 Related work

An alternative to directed graphical models with latent variables are undirected graphical models with latent variables, such as restricted Boltzmann machines (RBMs) [27, 16], deep Boltzmann machines (DBMs) [26] and their numerous variants. The interactions within such models are represented as the product of unnormalized potential functions, normalized by a global summa- tion/integration over all states of the random variables. This quantity (the partition function) and its gradient are intractable for all but the most trivial instances, although they can be estimated by Markov chain Monte Carlo (MCMC) methods. Mixing poses a significant problem for learning algorithms that rely on MCMC [3, 5].

Deep belief networks (DBNs) [16] are hybrid models containing a single undirected layer and sev- eral directed layers. While a fast approximate layer-wise training criterion exists, DBNs incur the computational difficulties associated with both undirected and directed models.

Alternative criteria that do not approximate or bound the log-likelihood have also been proposed, such as score matching [18] and noise-contrastive estimation (NCE) [13]. Both of these require the learned probability density to be analytically specified up to a normalization constant. Note that in many interesting generative models with several layers of latent variables (such as DBNs and DBMs), it is not even possible to derive a tractable unnormalized probability density. Some models such as denoising auto-encoders [30] and contractive autoencoders have learning rules very similar to score matching applied to RBMs. In NCE, as in this work, a discriminative training criterion is employed to fit a generative model. However, rather than fitting a separate discriminative model, the generative model itself is used to discriminate generated data from samples a fixed noise distribution.

Because NCE uses a fixed noise distribution, learning slows dramatically after the model has learned even an approximately correct distribution over a small subset of the observed variables.

Finally, some techniques do not involve defining a probability distribution explicitly, but rather train a generative machine to draw samples from the desired distribution. This approach has the advantage that such machines can be designed to be trained by back-propagation. Prominent recent work in this area includes the generative stochastic network (GSN) framework [5], which extends generalized denoising auto-encoders [4]: both can be seen as defining a parameterized Markov chain, i.e., one learns the parameters of a machine that performs one step of a generative Markov chain. Compared to GSNs, the adversarial nets framework does not require a Markov chain for sampling.

Because adversarial nets do not require feedback loops during generation, they are better able to leverage piecewise linear units [19, 9, 10], which improve the performance of backpropagation but have problems with unbounded activation when used ina feedback loop. More recent examples of training a generative machine by back-propagating into it include recent work on auto-encoding variational Bayes [20] and stochastic backpropagation [24].

3 Adversarial nets

The adversarial modeling framework is most straightforward to apply when the models are both multilayer perceptrons. To learn the generator's distribution pg over data x, we define a prior on input noise variables pz(z), then represent a mapping to data space as G(z; ´g), where G is a differentiable function represented by a multilayer perceptron with parameters ´g . We also define a second multilayer perceptron D(x; ´d) that outputs a single scalar. D(x) represents the probability that x came from the data rather than pg. We train D to maximize the probability of assigning the correct label to both training examples and samples from G.

In the next section, we present a theoretical analysis of adversarial nets, essentially showing that the training criterion allows one to recover the data generating distribution as G and D are given enough capacity, i.e., in the non-parametric limit. See Figure 1 for a less formal, more pedagogical explanation of the approach. In practice, we must implement the game using an iterative, numerical approach. Optimizing D to completion in the inner loop of training is computationally prohibitive, and on finite datasets would result in overfitting. Instead, we alternate between k steps of optimizing D and one step of optimizing G. This results in D being maintained near its optimal solution, so long as G changes slowly enough. This strategy is analogous to the way that SML/PCD [31, 29] training maintains samples from a Markov chain from one learning step to the next in order to avoid burning in a Markov chain as part of the inner loop of learning. The procedure is formally presented in Algorithm 1.
In practice, equation 1 may not provide sufficient gradient for G to learn well. Early in learning, when G is poor, D can reject samples with high confidence because they are clearly different from the training data. In this case, log(1 – D(G(z))) saturates. Rather than training G to minimize log(1 – D(G(z))) we can train G to maximize log D(G(z)). This objective function results in the same fixed point of the dynamics of G and D but provides much stronger gradients early in learning.

4 Theoretical Results

The generator G implicitly defines a probability distribution pg as the distribution of the samples G(z) obtained when z pz. Therefore, we would like Algorithm 1 to converge to a good estimator of pdata, if given enough capacity and training time. The results of this section are done in a non- parametric setting, e.g. we represent a model with infinite capacity by studying convergence in the space of probability density functions.

5 Experiments

We trained adversarial nets an a range of datasets including MNIST[23], the Toronto Face Database (TFD) [28], and CIFAR-10 [21]. The generator nets used a mixture of rectifier linear activations [19, 9] and sigmoid activations, while the discriminator net used maxout [10] activations. Dropout [17] was applied in training the discriminator net. While our theoretical framework permits the use of dropout and other noise at intermediate layers of the generator, we used noise as the input to only the bottommost layer of the generator network. We estimate probability of the test set data under pg by fitting a Gaussian Parzen window to the samples generated with G and reporting the log-likelihood under this distribution.

The reported numbers on MNIST are the mean log-likelihood of samples on test set, with the standard error of the mean computed across examples. On TFD, we computed the standard error across folds of the dataset, with a different ´ chosen using the validation set of each fold. On TFD, ´ was cross validated on each fold and mean log-likelihood on each fold were computed. For MNIST we compare against other models of the real-valued (rather than binary) version of dataset. of the Gaussians was obtained by cross validation on the validation set.

DISSIMILARITY CALCULATION

INFORMATION SPHERE | DISINFORMATION CAMPAIGN | ADVERSARIAL TRAINING | DISCRIMINATOR | NON-SENSICAL PATTERN | IS TRYING TO FOOL

lleux et al. [8] and
which the exact
ults are reported
likelihood has
erform well in high
thod available to
models than an
tly r

ages and
eling frame- works.
ere is no explicit
t be synchronized
G must not be
order to avoid "the
s too many values
gh diversity to
ins of a Boltzmann
en learning steps.
are never needed,
ts, no inference
riety of functions
le 2 summarizes
al nets with other

rarily
also gain some
r network not
ples, but only with
ator. This means
pied directly
r advantage of
hile methods
e distribution be
e to be able to m

rward extensions:

p(x | c) can be
and D.

Learned approximate inference can be performed by training an auxiliary network to predict z given x. This is similar to the inference net trained by the wake-sleep algorithm [15] but with the advantage that the inference net may be trained for a fixed generator net after the generator net has finished training.

3. One can approximately model all conditionals p(xS | x S) where S is a subset of the indices of x by training a family of conditional models that share parameters. Essentially, one can use adversarial nets to implement a stochastic extension of the deterministic MP-DBM [11].

Semi-supervised learning: features from the discriminator or inference net could improve performance of classifiers when limited label data is available.

5. Efficiency improvements: training could be accelerated greatly by divising better methods for coordinating G and D or determining better distributions to sample z from during training.

This paper has demonstrated the viability of the adversarial modeling framework, suggesting that these research directions could prove useful.


THE LATTER
ANDTHELATTERISTRYINGTO
A V O I D
BEING FOOLED
DEEP FAKE
SEMANTICALLY SIMILAR
MUTUAL INFORMATION

KINSHIP | RELATIONS

SPECIALLY CRAFTED

STRUCTURES OF

INTERCONNECTED

LAYERS | BLENDED FUTURE

SEMI-SUPERVISED

AUTONOMOUS AGENTS

Artificial intelligence and machine learning capabilities are growing
at an unprecedented rate. … beneficial applications,
ranging from machine tra… …ntless more such
applications are being de… …ng term. Less
attention has historically b… …ntelligence can be used
maliciously. This report su… …ty threats from malicious
…arning capabilities are
…y widely beneficial
…analysis. Countless more
…the long term. Less
…ntelligence can be used
…ty threats from malicious
…better forecast, prevent,
…solve, the question of
…s will be. We focus instead
…ely to see soon if adequate defenses are not developed.


STRATEGIC AFFAIRS
M A L I C I O U S
USE
We present a novel approach for real-time facial reenactment of a monocular target video sequence (e.g., Youtube video). The source sequence is also a monocular video stream, captured live with a commodity webcam. Our goal is to animate the facial expressions of the target video by a source actor and re-render the manipulated output video in a photo-realistic fashion. To this end, we first address the under-constrained problem of facial identity recovery from monocular video by non-rigid model-based bundling. At run time, we track facial expressions of both source and tar- get video using a dense photometric consistency measure. Reenactment is then achieved by fast and efficient deformation transfer between source and target. The mouth interior that best matches the retargeted expression is retrieved from the target sequence and warped to produce an accurate fit. Finally, we convincingly re-render the synthesized target face on top of the corresponding video stream such that it seamlessly blends with the real-world illumination. We demonstrate our method in a live setup, where Youtube videos are reenacted in real time.
R A N D O M
NOISE
COMPRESSED
R E P R E S E N T A T I O N

THISPERSON DOESNOTEXIST:

Martin Wühler --> Link

synthetic media or an object that never existed

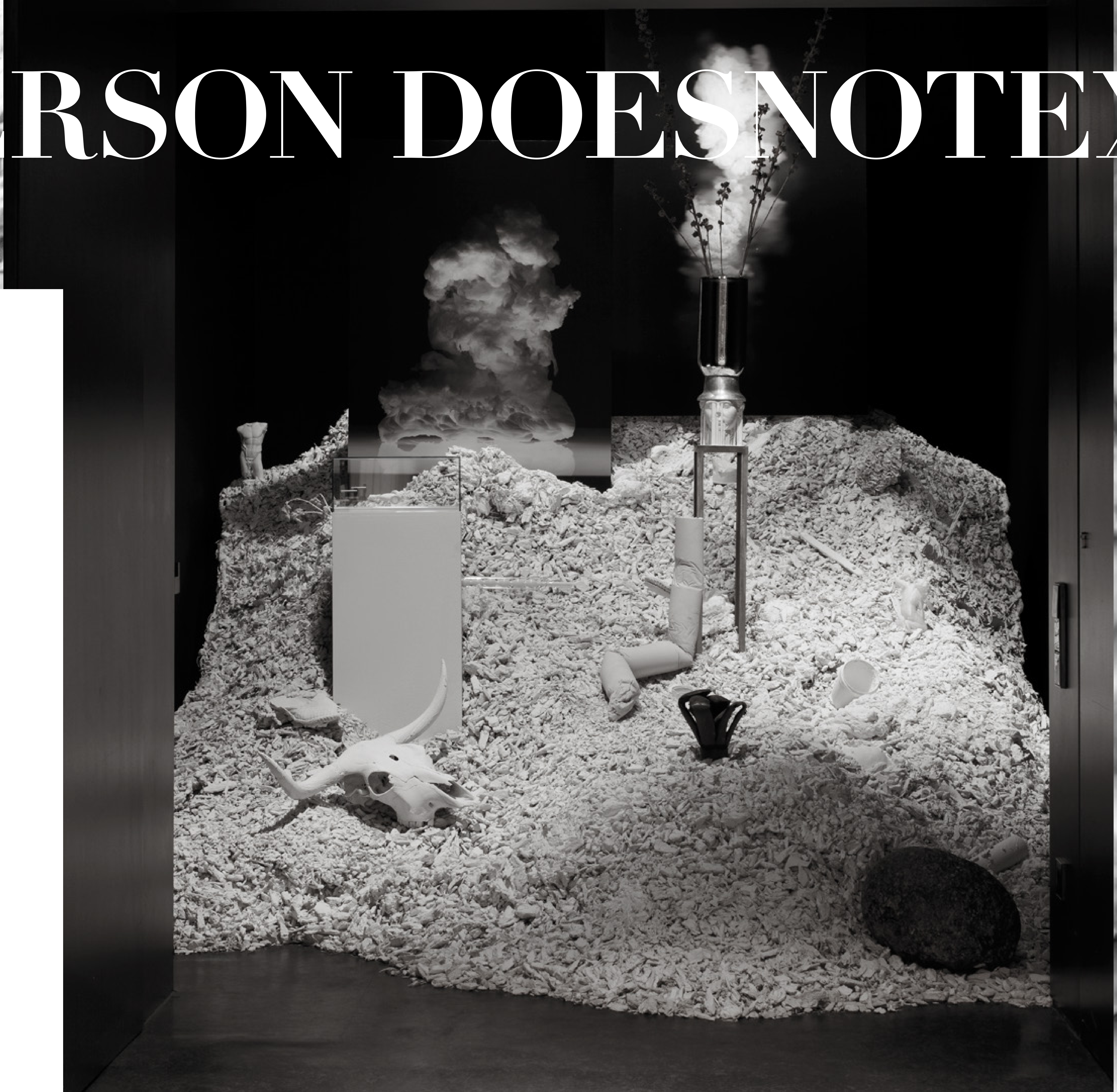
18.01. – 29.03.2020

45cbm --> Link

Staatliche Kunsthalle Baden-Baden --> Link

CURATOR

Hendrik Bündge



PERSONAL THANKS TO

the whole team of Staatliche Kunsthalle Baden-Baden

Freunde der Staatlichen Kunsthalle Baden-Baden E.V. --> Link

ANNALINDE gGmbH Leipzig --> Link

Justus Jager --> Link

Erik Swars --> Link

DESIGN / TYPESETTING

Martin Wühler

TYPE

Neue Haas Grotesk Display Pro

Didot

CONTACT --> mail