Generative Adversarial Networks

Ali Akbar Septiandri

Universitas Al Azhar Indonesia

January 5, 2019

Daftar isi

Bahan Bacaan

1. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y., 2014. Generative adversarial nets. In *NIPS* (pp. 2672-2680).
2. Karpathy, et al. 16 June 2016. Generative Models. [Online] URL: `https://blog.openai.com/generative-models/`
3. Goodfellow, I., 2016. NIPS 2016 tutorial: Generative adversarial networks. arXiv preprint arXiv:1701.00160.
4. Carey, O. 14 September 2018. Generative Adversarial Networks (GANs) - A Beginner's Guide. [Online] URL: `https://towardsdatascience.com/generative-adversarial-networks-gans-a-beginners-guide-5b`

Model Generatif

Model Generatif

- Variational Autoencoder (VAE) dapat menghasilkan data baru
- Asumsi mengikuti distribusi normal
- “Kode” sebagai parameter distribusi
- Data baru adalah hasil *sampling* dari distribusi

Alternatif

- Alternatif untuk VAE: Generative Adversarial Networks (GANs) [Goodfellow et al., 2014]
- "This, and the variations that are now being proposed is the most interesting idea in the last 10 years in ML, in my opinion." (LeCun, 2016)
- Juga menggunakan asumsi distribusi sederhana, e.g. Gaussian, dan *sampling*

Latar Belakang

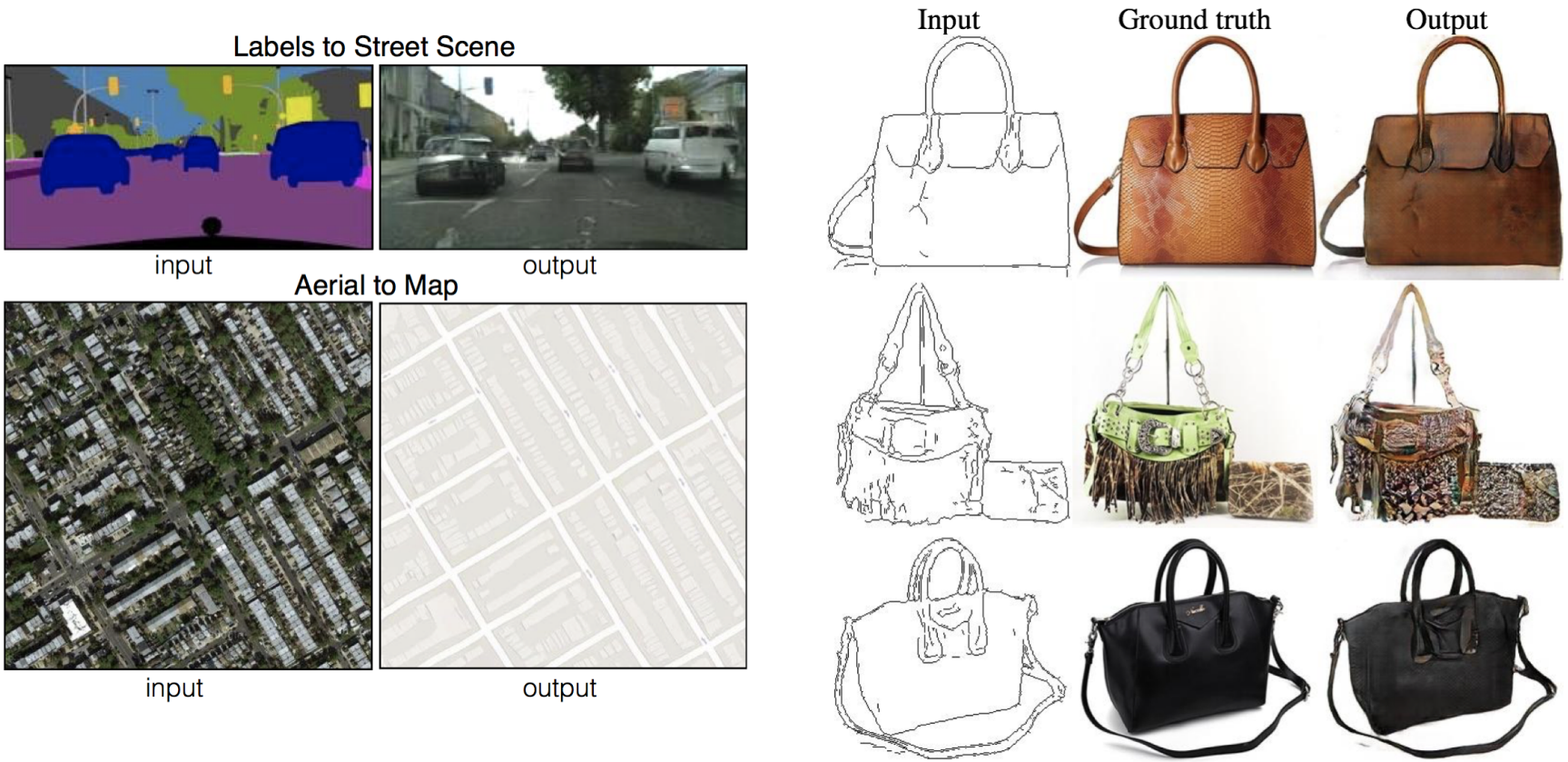
Dari [Goodfellow, 2016]:

- Melatih dan sampling dari model generatif membantu menguji kemampuan representasi dan manipulasi distribusi dalam dimensi tinggi
- Dapat dimasukkan dalam *reinforcement learning*
- Menangani kasus data yang hilang

Kasus yang Membutuhkan Sampel

- *Single image super-resolution*
- Menghasilkan karya seni — dapat terjual hingga $432,500!
- *Image-to-image translation*, e.g. sketsa menjadi gambar

Image-to-image Translation



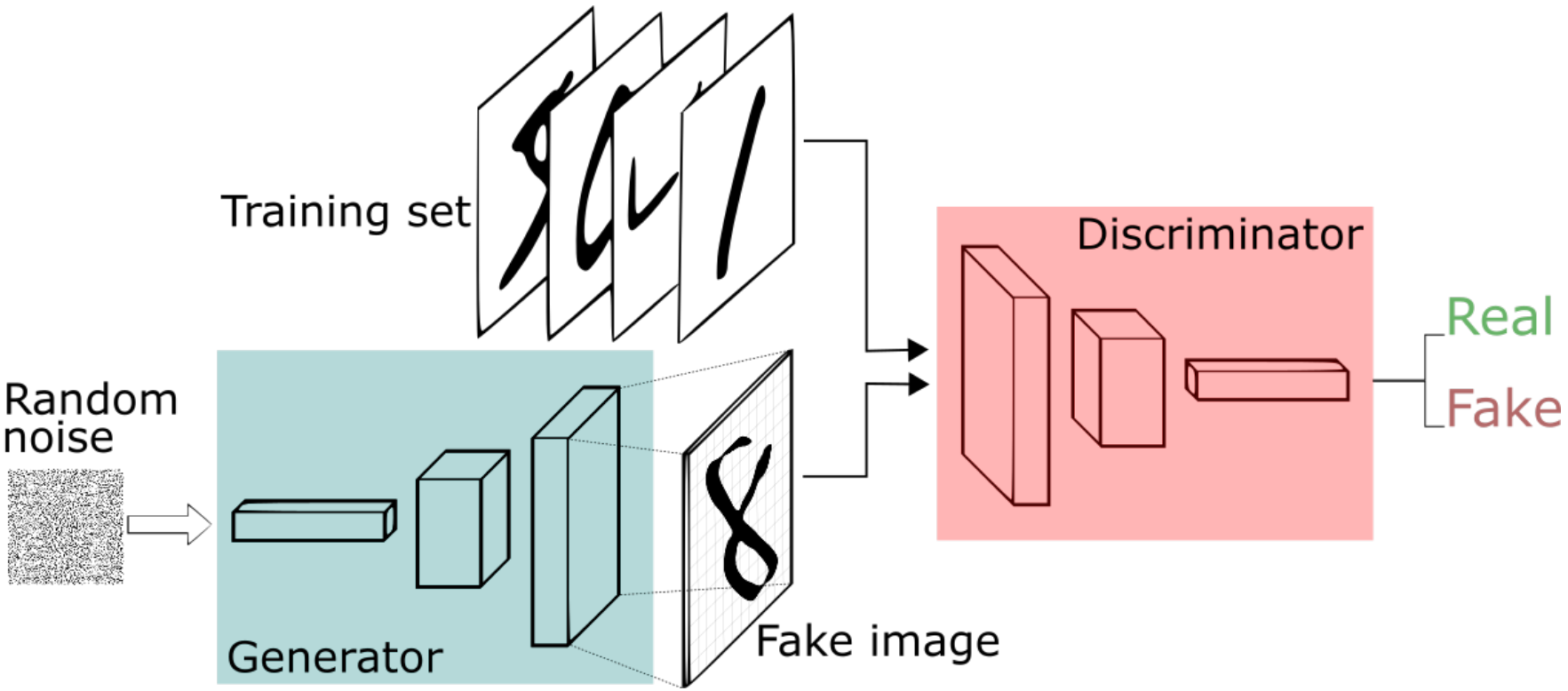
Gambar: Dari sketsa ke gambar atau sebaliknya (Isola et al., 2016)

Generative Adversarial Networks

Cara Kerja

- Melatih dua *networks*: generator $G(\mathbf{z}; \theta_g)$ dan diskriminator $D(\mathbf{x}; \theta_d)$
- Analogi:
  - generator = pemalsu
  - diskriminator = detektif
- Minimax game!

Ilustrasi



Gambar: Arsitektur GANs

Loss Functions

$$J^{(D)} = -\frac{1}{2}\mathbb{E}_{\mathbf{x}\sim p_{data}} \log D(\mathbf{x}) - \frac{1}{2}\mathbb{E}_{\mathbf{z}} \log(1 - D(G(\mathbf{z})))$$

$$J^{(G)} = -J^{(D)}$$

Pada praktiknya, gradien dari loss function untuk generator cepat jenuh.
Perlu pendekatan heuristik.

Perubahan Loss Functions

$$J^{(D)} = -\frac{1}{2}\mathbb{E}_{\mathbf{x}\sim p_{data}} \log D(\mathbf{x}) - \frac{1}{2}\mathbb{E}_{\mathbf{z}} \log(1 - D(G(\mathbf{z})))$$

$$J^{(G)} = -\frac{1}{2}\mathbb{E}_{\mathbf{z}} \log D(G(\mathbf{z}))$$

Generator berusaha memaksimalkan kesalahan diskriminator saja

Perbandingan

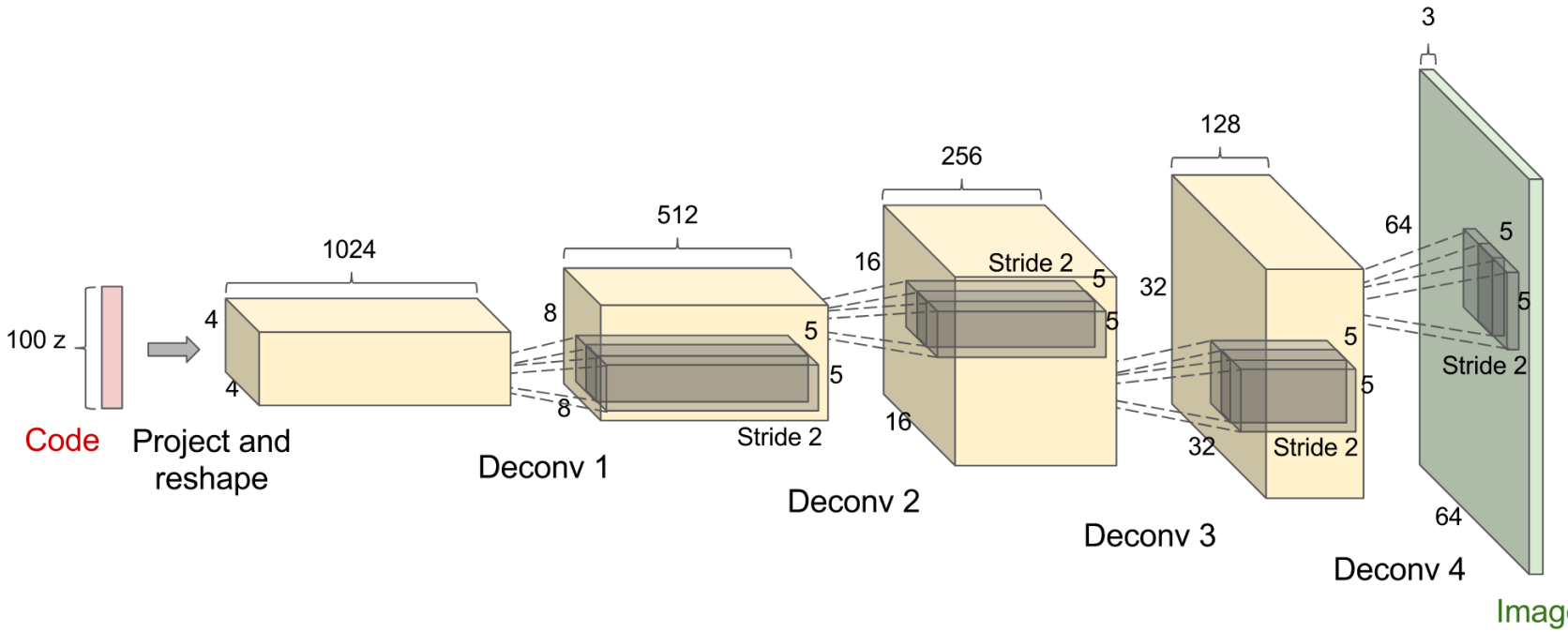
- Menggunakan "kode" laten
- *Asymptotically consistent*, i.e. jika diberikan data sangat banyak, semua hasilnya bisa diaproksimasi
- Tidak perlu Markov chain
- Sering dianggap menghasilkan data yang bagus

Tips

- Normalisasi input
- BatchNorm
- LeakyReLU
- Soft and Noisy Labels
- DCGAN
- dll. [Chintala et al., 2016]

Perkembangan

DCGAN

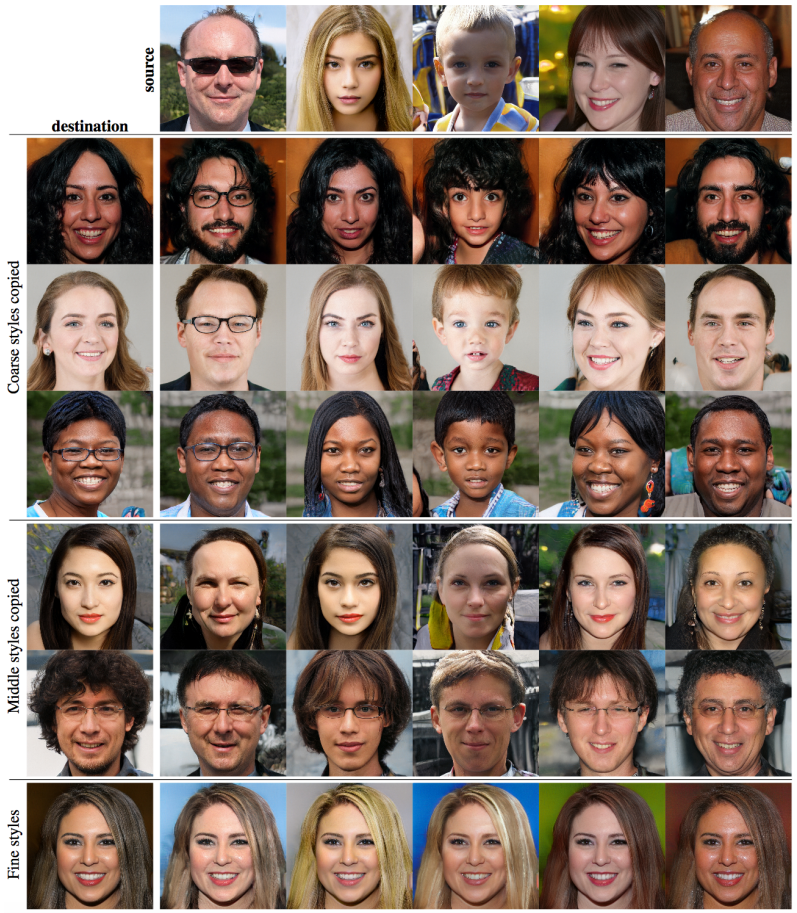


Gambar: DCGAN dengan BatchNorm, all-convolutional, dan Adam (Radford et al., 2015)

BigGAN

Gambar: Pengembangan GAN dengan data yang sangat banyak menghasilkan citra yang sangat riil [Brock et al., 2018]

Style-Based Generator for GANs



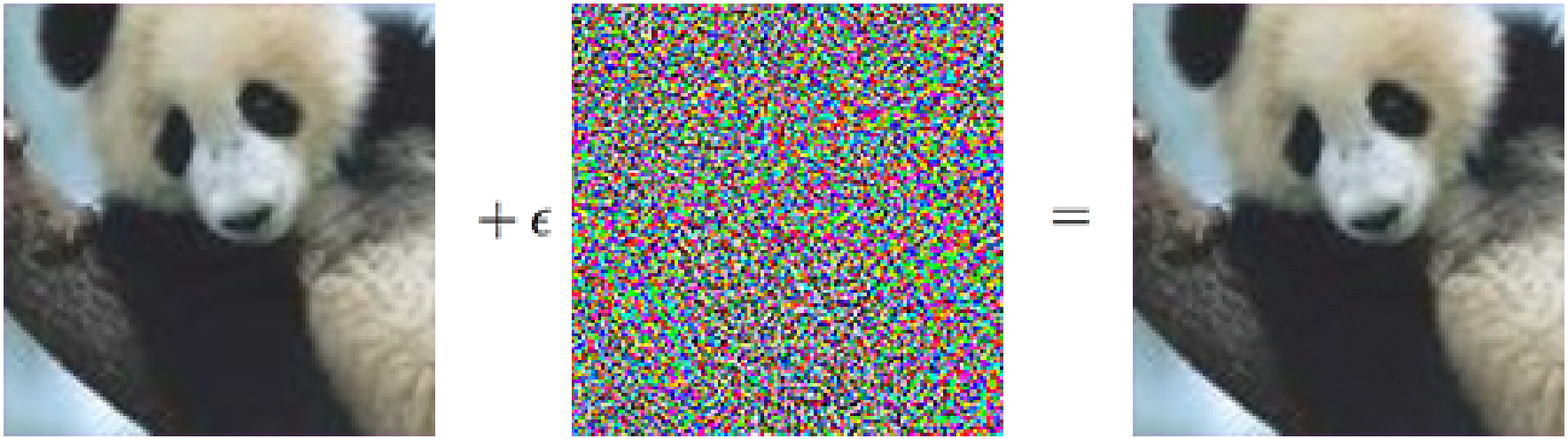
Gambar: Memanfaatkan ide *style transfer* untuk GANs [Karras et al., 2018]

Adversarial Attacks

Adversarial Examples

- ML classifiers mudah "tertipu" dengan *adversarial examples*
- Ancaman di bidang keamanan informasi
- Gambar yang telah dicetak bahkan tetap bisa salah diklasifikasi [Kurakin et al., 2017]

Contoh



Gambar: Kesalahan klasifikasi hanya dengan menambahkan *noise*

"Adversarial examples are hard to defend against because it is difficult to construct a theoretical model of the adversarial example crafting process." [Goodfellow et al., 2017]

Ikhtisar

1. GAN adalah model generatif yang dapat mensimulasi berbagai *cost function*
2. Dapat dilihat sebagai game dengan dua pemain, diskriminator dan generator, yang saling berlawanan
3. Merupakan bidang yang sangat berkembang dan dapat digunakan untuk berbagai kasus

Referensi

Ian Goodfellow et al. (2014)

Generative Adversarial Nets

In *NIPS* (pp. 2672-2680)

Karpathy et al. (June 2016)

Generative Models

https://blog.openai.com/generative-models/

Ian Goodfellow (2016)

NIPS 2016 Tutorial: Generative Adversarial Networks

https://arxiv.org/pdf/1701.00160.pdf

Soumith Chintala et al. (2016)

How to Train a GAN? Tips and tricks to make GANs work

https://github.com/soumith/ganhacks

Referensi

Andrew Brock, Jeff Donahue, & Karen Simonyan (2018)

Large Scale GAN Training for High Fidelity Natural Image Synthesis

https://arxiv.org/pdf/1809.11096.pdf

Tero Karras, Samuli Laine, & Timo Aila (2018)

A Style-Based Generator Architecture for Generative Adversarial Networks

https://arxiv.org/pdf/1812.04948.pdf

Alexey Kurakin, Ian Goodfellow, & Samy Bengio (2017)

Adversarial Examples in the Physical World

https://openreview.net/pdf?id=S1OufnIlx

Ian Goodfellow et al. (2017)

Attacking Machine Learning with Adversarial Examples

https://blog.openai.com/adversarial-example-research/

Terima kasih