

Manipulasi Gambar Menggunakan Jaringan *Adversarial* Siklus-Konsisten

Image Manipulation Using a Cycle-Consistent Adversarial Network

Afrizal Zein

Program Studi Teknik Informatika, STMIK Eresha
 Jl. Raya Puspitek Serpong No. 10 Tangerang Selatan Banten
 E-mail : zeinafrizal@gmail.com

Abstrak--- *Tahukah Anda bahwa memanipulasi gambar merupakan hal yang menarik dan disenangi orang, karena dapat menampilkan hal yang baru lain daripada yang lain. Umumnya manipulasi gambar kita menggunakan software aplikasi seperti photoshop dan banyak lagi software untuk hal tersebut. Penelitian ini mencoba memanipulasi gambar dengan teori yang ada dan diaplikasikan dengan membuat program dalam bahasa python. Adapun metoda yang digunakan untuk penelitian ini adalah metoda jaringan adversarial siklus-konsisten yaitu mempelajari pemetaan antara gambar input dan gambar output menggunakan set pelatihan pasangan gambar yang selaras. Kami menyajikan pendekatan untuk belajar menerjemahkan gambar dari domain sumber X ke domain target Y dengan tidak adanya contoh berpasangan. Tujuan kami adalah mempelajari pemetaan $G: X \rightarrow Y$ sedemikian rupa sehingga distribusi gambar dari $G(X)$ tidak dapat dibedakan. Hasil manipulasi gambar yang dihasilkan sangatlah sempurna menyerupai gambar asli tapi berbeda tampilan.*

Kata Kunci : *Manipulasi Gambar, Jaringan Adversarial Siklus-Konsisten, Deeplearning*

Abstract--- *Did you know that manipulating images is an interesting thing and is liked by people, because it can show new things than others. Generally, image manipulation, we use application software such as Photoshop and many other software for that. This research tries to manipulate images with existing theories and is applied by making programs in python language. The method used for this study is a cycle-consistent adversarial network method, namely studying the mapping between input images and output images using a training set of matching pairs of images. We present an approach to learning to translate images from source X domains to target Y domains in the absence of paired examples. Our aim is to study the mapping of $G: X \rightarrow Y$ in such a way that the distribution of images from $G(X)$ cannot be distinguished. The results of manipulation of the images produced are very perfect resembling the original image but different appearance.*

Keywords : *Image Manipulation, Cycle-Consistent Adversarial Network, Deeplearning*

1. PENDAHULUAN

Manipulasi Citra (gambar) adalah proses melalui tampilan visual dari gambar asli yang telah diubah. Teknik-teknik ini dapat digunakan untuk memproduksi efek yang lebih luas, dari perubahan halus dan koreksi untuk lebih intervensi dramatis, obyektif, dapat diandalkan, dan efisien untuk melakukannya.

Untuk saat ini penggunaan dari manipulasi citra telah melingkupi berbagai macam disiplin ilmu diantaranya bidang arsitektur, geografi, ilmu komputer, kedokteran, fotografi, arkeologi, dan lain sebagainya.

Tujuan utama dari digital image processing tersebut selain untuk meningkatkan kualitas gambar yang diperoleh, juga dimaksudkan untuk memproses data yang diperoleh untuk ditanggapi secara otomatis oleh sebuah mesin atau peralatan.

Masalah ini dapat secara lebih luas digambarkan sebagai imajinasi terjemahan gambar, mengkonversi gambar dari satu representasi adegan

tertentu, x , ke adegan lain, y , misal, skala abu-abu ke warna, gambar ke label semantik, tepi-peta ke foto. Tahun penelitian dalam visi komputer, gambar pemrosesan, fotografi komputasi, dan grafis memiliki menghasilkan sistem terjemahan yang kuat di bawah pengawasan pengaturan, di mana contoh pasangan gambar $f(x_i); y(i) = 1$ tersedia.

Namun, memperoleh data pelatihan berpasangan mungkin sulit dan mahal. Misalnya, hanya beberapa set data yang ada untuk tugas-tugas seperti segmentasi semantik dan mereka relatif kecil. Memperoleh pasangan input-output untuk grafik tugas-tugas seperti gaya artistik bisa lebih sulit karena output yang diinginkan sangat kompleks, biasanya membutuhkan penulisan artistik. Untuk banyak tugas, seperti mentranfer konfigurasi objek misal, zebra dan kuda, yang diinginkan output bahkan tidak didefinisikan dengan baik.

Karena itu kami mencari algoritma yang dapat belajar menerjemahkan antar domain tanpa

contoh hasil masukan dan keluaran yang berpasangan. Kami berasumsi ada beberapa hubungan yang mendasarinya antara domain misalnya, apakah itu dua rendering berbeda dari adegan dasar yang sama dan berusaha mempelajari hubungan itu. Meski kami kurang pengawasan dalam bentuk contoh berpasangan, kita dapat memanfaatkan pengawasan di tingkat set: kami diberi satu set gambar di domain X dan set berbeda di domain Y . Kami mungkin berlatih di sebuah pemetaan $G: X \rightarrow Y$ sedemikian rupa sehingga keluaran $\hat{y} = G(x)$, $x \in X$, tidak dapat dibedakan dari gambar $y \in Y$ oleh musuh dilatih untuk mengklasifikasikan \hat{y} terpisah dari y . Secara teori, tujuan ini dapat menginduksi distribusi output lebih dari \hat{y} yang cocok pada distribusi empiris (y) (secara umum, ini membutuhkan G menjadi stochastic) [16]. G yang optimal diterjemahkan domain X ke domain Y didistribusikan secara identik ke Y . Namun, terjemahan seperti itu tidak menjamin bahwa seorang individu input x dan output y dipasangkan secara bermakna - ada banyak pemetaan G yang akan mendorong distribusi yang sama lebih dari \hat{y} . Apalagi dalam praktiknya, kami merasa sulit untuk mengoptimalkan tujuan per musuh dalam isolasi: prosedur standar sering mengarah pada yang terkenal masalah mode runtuh, di mana semua input gambar memetakan ke gambar output yang sama dan optimasi gagal membuat kemajuan.

Masalah-masalah ini membutuhkan tambahan struktur untuk tujuan kami. Karena itu, kami memanfaatkan properti terjemahan itu harus "konsisten siklus", dalam arti bahwa jika kita menerjemahkan, misal, kalimat dari Bahasa Inggris ke Bahasa Prancis, lalu terjemahkan kembali dari Prancis ke Inggris, kita harus kembali ke kalimat aslinya [3]. Secara matematis, jika kita memiliki a penerjemah $G: X \rightarrow Y$ dan penerjemah lain $F: Y \rightarrow X$, maka G dan F harus merupakan invers satu sama lain, dan keduanya pemetaan harus menjadi bijections. Kami menerapkan asumsi struktural ini dengan melatih pemetaan G dan F secara bersamaan, dan menambahkan kehilangan konsistensi siklus [64] yang mendorong $F(G(x)) \approx x$ dan $G(F(y)) \approx y$. Menggabungkan kehilangan ini dengan kerugian permusuhan pada domain X dan Y menghasilkan penuh kami obyektif untuk terjemahan gambar ke gambar yang tidak berpasangan. Kami menerapkan metode kami untuk berbagai aplikasi, termasuk transfer gaya koleksi, transfigurasi objek, transfer musim dan peningkatan foto. Kami juga membandingkan terhadap pendekatan sebelumnya yang bergantung pada definisi tangan faktorisasi gaya dan konten, atau pada *embedding* bersama berfungsi, dan menunjukkan bahwa metode kami mengungguli ini garis dasar.

2. TINJAUAN PUSTAKA

Menurut penelitian yang dilakukan oleh Ming-Yu Liu, Thomas Breuel, Jan Kautz (2018) yang berjudul "*Unsupervised Image-to-Image Translation Networks*", latar belakang dan tujuan: Terjemahan gambar ke gambar tanpa pengawasan bertujuan untuk mempelajari distribusi bersama dari gambar di domain yang berbeda dengan menggunakan gambar dari distribusi marginal di domain individual. Karena ada set distribusi bersama yang tak terbatas itu dapat tiba distribusi marginal yang diberikan, orang bisa menyimpulkan apa-apa tentang sambungan distribusi dari distribusi marginal tanpa asumsi tambahan. Untuk mengatasi masalah tersebut, kami membuat asumsi ruang bersama-laten dan mengusulkan kerangka terjemahan gambar ke gambar tanpa pengawasan berdasarkan GAN yang digabungkan.

Kami membandingkan kerangka kerja yang diusulkan dengan pendekatan yang bersaing dan hadir hasil terjemahan gambar berkualitas tinggi pada berbagai gambar tanpa pengawasan yang menantang tugas terjemahan, termasuk terjemahan gambar pemandangan jalan, terjemahan gambar hewan, dan terjemahan gambar wajah. Kami juga menerapkan kerangka kerja yang diusulkan ke domain adaptasi dan mencapai kinerja canggih pada dataset benchmark.

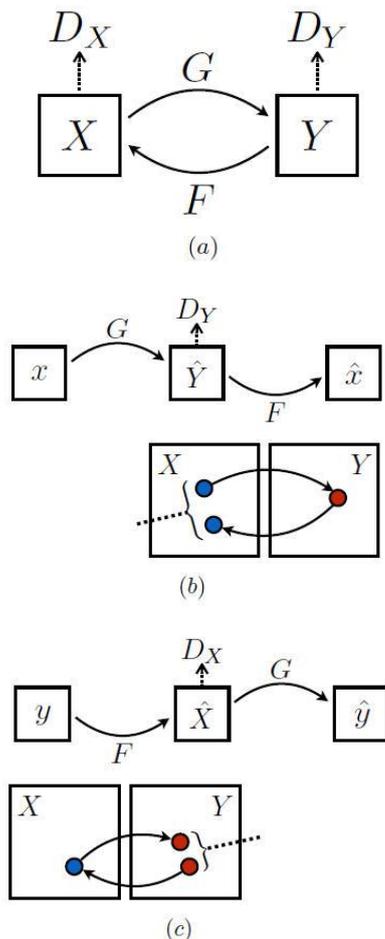
Menurut penelitian yang dilakukan oleh Phillip Isola Jun-Yan Zhu Tinghui Zhou Alexei A. Efros (2017), yang berjudul "*Image-to-Image Translation with Conditional Adversarial Networks*", kami menyelidiki jaringan bersyarat sebagai sebuah solusi tujuan umum untuk terjemahan gambar ke gambar. Jaringan ini tidak hanya mempelajari pemetaan dari input gambar ke gambar output, tetapi juga belajar fungsi kerugian untuk melatih pemetaan ini. Ini memungkinkan untuk diterapkan pendekatan generik yang sama untuk masalah yang secara tradisional akan membutuhkan formulasi kerugian yang sangat berbeda. Kami berdemonstrasi bahwa pendekatan ini efektif dalam mensintesis foto dari label-label, merekonstruksi objek dari peta tepi, dan mewarnai gambar, di antara tugas-tugas lainnya. Memang sejak itu rilis perangkat lunak pix2pix yang terkait dengan makalah ini, sejumlah besar pengguna internet (banyak dari mereka artis) telah memposting eksperimen mereka sendiri dengan sistem kami, lebih lanjut menunjukkan penerapannya yang luas dan kemudahan adopsi tanpa perlu tweak parameter. Sebagai sebuah komunitas, kami tidak lagi merekayasa fungsi pemetaan kami, dan karya ini menyarankan agar kita dapat mencapai hasil yang masuk akal tanpa rekayasa tangan fungsi kerugian kita juga.

3. METODA

Ide dari aplikasi ini adalah memanipulasi dari gambar ke gambar yang menggunakan tekstur non-parametrik

Model pada pasangan gambar pelatihan input-output tunggal. Pendekatan yang lebih baru menggunakan dataset contoh input-output untuk mempelajari fungsi terjemahan parametrik menggunakan *Convolution Neural Network*.

Pendekatan kami dibangun di atas kerangka "pix2pix" yang menggunakan generatif bersyarat jaringan permusuhan untuk mempelajari pemetaan dari input ke gambar output. Ide serupa telah diterapkan ke berbagai tugas seperti menghasilkan foto dari sketsa atau dari atribut dan tata letak semantik]. Namun tidak seperti itu pekerjaan sebelumnya di atas, kita belajar pemetaan tanpa dipasangkan.



Gambar 1. (a) Model kami berisi dua fungsi pemetaan $G: X \rightarrow Y$ dan $F: Y \rightarrow X$, dan permusuhan terkait diskriminator D_Y dan D_X . D_Y mendorong G untuk menerjemahkan X ke dalam keluaran yang tidak dapat dibedakan dari domain Y , dan sebaliknya untuk D_X dan F . Untuk lebih jauh mengatur pemetaan, kami memperkenalkan dua kerugian konsistensi siklus yang menangkap intuisi bahwa jika kita menerjemahkan dari satu domain ke domain

lain dan kembali lagi kita harus sampai di tempat kita mulai: (b) meneruskan konsistensi siklus kerugian: $x! G(x)! F(G(x))x$, dan (c) kehilangan konsistensi siklus mundur: $y! F(y)! G(F(y))y$

Terjemahan Gambar ke Gambar yang Tidak Berpasangan

Beberapa lainnya metode juga menangani pengaturan tidak berpasangan, di mana tujuannya untuk menghubungkan dua domain data: X dan Y . Mengusulkan kerangka kerja Bayesian yang mencakup berbasis sebelumnya pada bidang acak Markov berbasis *patch* yang dihitung dari sebuah gambar sumber dan istilah kemungkinan diperoleh dari banyak gambar gaya. Baru-baru ini, *CoGAN* menggunakan strategi berbagi berat untuk belajar sebuah representasi umum di seluruh domain. Bersamaan dengan kami metode, Liu memperluas kerangka kerja di atas dengan kombinasi variabel *autoencoder* dan generative jaringan permusuhan. Garis lain bersamaan mendorong input dan output untuk dibagikan fitur "konten" tertentu meskipun mungkin berbeda "gaya". Metode-metode ini juga menggunakan jaringan permusuhan, dengan persyaratan tambahan untuk menegakkan output agar dekat dengan input dalam ruang metrik yang telah ditentukan, seperti ruang label kelas, ruang piksel, dan ruang fitur.

Berbeda dengan pendekatan di atas, formulasi kami tidak bergantung pada fungsi kesamaan yang ditentukan oleh tugas apa pun yang spesifik input dan output, kami juga tidak menganggap bahwa input dan output harus terletak pada *embedding* dimensi rendah yang sama ruang. Ini menjadikan metode kami sebagai solusi tujuan umum untuk banyak tugas visi dan grafik. Kami langsung membandingkan terhadap beberapa pendekatan sebelumnya dan kontemporer di Indonesia.

Konsistensi Siklus

Gagasan untuk menggunakan transitivitas sebagai sebuah cara untuk mengatur data terstruktur memiliki sejarah panjang. Di pelacakan visual, menegakkan konsistensi adalah hal sederhana yang telah menjadi trik standar selama beberapa dekade.

Di domain bahasa, memverifikasi dan meningkatkan terjemahan via "terjemahan balik dan rekonsiliasi" adalah teknik digunakan oleh penerjemah manusia, serta oleh mesin. Lebih baru-baru ini, konsistensi siklus tingkat tinggi telah digunakan di struktur dari gerak, pencocokan bentuk 3D, cosegmentasi, perataan semantik yang padat, dan estimasi kedalaman. Dari jumlah tersebut, paling mirip dengan pekerjaan kami, karena mereka menggunakan sebuah kehilangan konsistensi siklus sebagai cara menggunakan transitivitas untuk mengawasi Pelatihan CNN. Dalam karya ini, kami memperkenalkan sebuah

kerugian serupa untuk mendorong G dan F agar konsisten dengan masing-masing lain. Bersamaan dengan pekerjaan kami, dalam proses yang sama ini, secara independen menggunakan tujuan yang sama untuk terjemahan gambar ke gambar yang tidak berpasangan, terinspirasi oleh dual belajar dalam terjemahan mesin.

4. HASIL DAN PEMBAHASAN

Detail pelatihan Kami menerapkan dua teknik dari yang terbaru bekerja untuk menstabilkan prosedur pelatihan model kami.

$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{x,y}[\log D(x, y)] + \mathbb{E}_{x,z}[\log(1 - D(x, G(x, z)))], \quad (1)$$

Pertama, untuk LGAN (Persamaan 1), kami mengganti kemungkinan log negative objektif dengan kerugian kuadrat terkecil. Kerugian ini lebih stabil selama pelatihan dan menghasilkan kualitas yang lebih tinggi hasil.

$$\mathcal{L}_{GAN}(G, D) = \mathbb{E}_y[\log D(y)] + \mathbb{E}_{x,z}[\log(1 - D(G(x, z)))]. \quad (2)$$

Khususnya, untuk kerugian GAN LGAN (G; D; X; Y), kami melatih G untuk meminimalkan $\mathbb{E}_x p_{data}(x) [(D(G(x)) - 1)^2]$ dan latih D untuk meminimalkan $\mathbb{E}_y p_{data}(y) [(D(y) - 1)^2] + \mathbb{E}_x p_{data}(x) [D(G(x)) - 1]^2$. Kedua, untuk mengurangi osilasi model, kami ikuti strategi Shrivastava dan memperbarui diskriminator menggunakan riwayat gambar yang dihasilkan daripada yang diproduksi oleh generator terbaru. Kami menyimpan gambar buffer yang menyimpan 50 gambar yang dibuat sebelumnya.

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y,z}[\|y - G(x, z)\|_1]. \quad (3)$$

Untuk semua percobaan, kami menetapkan = 10 dalam Persamaan 3. Kami menggunakan pemecah Adam [26] dengan ukuran *batch* 1. Semua jaringan dilatih dari awal dengan tingkat pembelajaran 0 : 0002. Kami mempertahankan tingkat pembelajaran yang sama untuk 100 pertama zaman dan secara linear mengurangi tingkat ke nol selama 100 berikutnya zaman. Silakan lihat lampiran (Bagian 7) untuk lebih jelasnya tentang dataset, arsitektur, dan prosedur pelatihan.

```
import time
from options.train_options import TrainOptions
from data import create_dataset
from models import create_model
from util.visualizer import Visualizer

if __name__ == '__main__':
    opt = TrainOptions().parse() # opsi pelatihan
    dataset = create_dataset(opt) # create a dataset
    dataset_size = len(dataset)
    print("The number of training images = %d" %
          dataset_size)

    model = create_model(opt) # create model
```

```
setup(opt)
visualizer = Visualizer(opt)
total_iters = 0
for epoch in range(opt.epoch_count, opt.niter +
opt.niter_decay + 1): # outer loop for different
epochs; we save the model by <epoch_count>,
<epoch_count>+<save_latest_freq>
    epoch_start_time = time.time()
    iter_data_time = time.time()
    epoch_iter = 0
epoch, reset to 0 every epoch
for i, data in enumerate(dataset):
    iter_start_time = time.time()
    if total_iters % opt.print_freq == 0:
        t_data = iter_start_time - iter_data_time
        visualizer.reset()
        total_iters += opt.batch_size
        epoch_iter += opt.batch_size
        model.set_input(data)
        model.optimize_parameters()
        if total_iters % opt.display_freq == 0: #
display images on visdom and save images to a
HTML file
            save_result = total_iters %
opt.update_html_freq == 0
            model.compute_visuals()

visualizer.display_current_results(model.get_curre
nt_visuals(), epoch, save_result)

        if total_iters % opt.print_freq == 0:
            losses = model.get_current_losses()
            t_comp = (time.time() - iter_start_time) /
opt.batch_size
            visualizer.print_current_losses(epoch,
epoch_iter, losses, t_comp, t_data)
            if opt.display_id > 0:
                visualizer.plot_current_losses(epoch,
float(epoch_iter) / dataset_size, losses)

        if total_iters % opt.save_latest_freq == 0:

            print('saving the latest model (epoch %d,
total_iters %d)' % (epoch, total_iters))
            save_suffix = 'iter_%d' % total_iters if
opt.save_by_iter else 'latest'
            model.save_networks(save_suffix)

            iter_data_time = time.time()
            if epoch % opt.save_epoch_freq == 0:
                print('saving the model at the end of epoch
%d, iters %d' % (epoch, total_iters))
                model.save_networks('latest')
                model.save_networks(epoch)

            print('End of epoch %d / %d \t Time Taken:
%d sec' % (epoch, opt.niter + opt.niter_decay,
time.time() - epoch_start_time))
            model.update_learning_rate()
```

Kami pertama kali membandingkan pendekatan kami terhadap metode terbaru untuk terjemahan gambar ke gambar yang tidak berpasangan pada kumpulan data yang berpasangan di mana pasangan input output ground-ground tersedia untuk evaluasi.

Kami kemudian mempelajari pentingnya kedua permusuhan tersebut kehilangan dan siklus kehilangan konsistensi dan membandingkan penuh kami metode terhadap beberapa varian. Akhirnya, kami menunjukkan generalisasi algoritma kami pada berbagai aplikasi di mana data berpasangan tidak ada.

Untuk menjalankan program

```
# python test.py --dataroot ./datasets/facades --
name facades_pix2pix --model pix2pix --direction
BtoA
```

Hasil yang keluar :



Pohon Apel Ke Pohon Jeruk



Pohon Jeruk Ke Pohon Apel



5. KESIMPULAN

1. Meskipun metode kami dapat mencapai hasil yang menarik dibanyak kasus, hasilnya jauh dari positif seragam. Angka 17 menunjukkan beberapa kasus kegagalan yang khas.
2. Kita punya juga mengeksplorasi tugas yang memerlukan perubahan geometris, dengan sedikit keberhasilan. Misalnya, pada tugas transfigurasi Anjing ke Kucing, terjemahan yang dipelajari berubah menjadi minimal. Kegagalan ini mungkin disebabkan oleh arsitektur generator kami yang dirancang khusus untuk performa bagus pada perubahan penampilan.

DAFTAR PUSTAKA

- [1] R. Rosales, K. Achan, and B. J. Frey. "Unsupervised Image Translation", In ICCV, 2003.
- [2] M.-Y. Liu, T. Breuel, and J. Kautz. "Unsupervised Image-To-Image Translation Networks". In NIPS, 2017.
- [3] L. A. Gatys, A. S. Ecker, and M. Bethge. "Image Style Transfer Using Convolutional Neural Networks". CVPR, 2016.
- [4] Afrizal Zein (2016), "Pendeteksian Multi Wajah dan Recognition Secara Real Time Menggunakan Metoda Principal Component Analisis", Jurnal ESIT Vol 12 No. 1 (2018).
- [5] Afrizal Zein (2018), "Menggabungkan Dua Wajah Dengan Metoda Ensemble Regression Trees Menggunakan Pustaka Dlib Dan Opencv Python", Jurnal ESIT Vol 13 No. 2 (2018).
- [6] Afrizal Zein (2018), "Pendeteksian Kantuk Secara Real Time Menggunakan Pustaka OPENCV dan DLIB PYTHON", Jurnal SAINSTECH Vol. 28 No. 2, (2018).