Translasi Citra Antara Manusia Dan Wayang Orang Menggunakan *Generative Adversarial Network*

Ciara Nurdenara[1], Wikky Fawwaz[2]

[1,2] Universitas Telkom, Bandung
[1]ciarand@students.telkomuniversity.ac.id, [2]wikkyfawwaz@telkomuniversity.ac.id

Abstrak

Hanya sedikit orang yang mengenal nudaya tradisional Indonesia, khususnya pertunjukkan wayang orang. Pemain wayang orang membutuhkan waktu sekitar sejam untuk menjadi wayang orang. Perlu menyewa kostum unik yang sulit ditemukan dan perlu mempelajari tata rias wayang orang yang bisa memakan waktu lama untuk menguasai. Penggunaan translasi citra dapat memudahkan setiap orang berkesempatan melihat diri berwujud wayang orang. Penelitian ini bertujuan untuk menerjemahkan wajah manusia menjadi wayang orang dengan menambahkan makeup dan aksesoris menggunakan *Generative Adversarial Network* (GAN) dan menggunakan *unpaired dataset* yang terdiri dari 1216 data latih dan 240 data uji. Tantangan dari penelitian ini adalah mempertahankan latar belakang citra dan komponen identitas wajah pada citra masukan. Penelitian ini menggunakan pengujian kuantitatif menggunakan FID, KID, dan IS untuk mengevaluasi kualitas citra yang dihasilkan dari generator. Hasil eksperimen dari penelitian ini adalah UGATIT memiliki hasil yang lebih baik dari DCLGAN berdasarkan nilai dari *Inception Score*, FID, dan KID. Berdasarkan IS, FID, dan KID, UGATIT memiliki hasil yang lebih baik daripada DCLGAN. Hasil dari UGATIT pada IS, FID, dan KID dengan skor sebagai berikut adalah 2.414, 0.924, dan 4.357, yang berarti UGATIT dapat berkinerja lebih baik daripada DCLGAN.

Kata kunci : translasi citra, GAN, *unpaired dataset,* wayang orang.

Abstract

Only a few people know about Indonesian traditional culture, especially wayang orang performances. Thewayang orang players took about an hour to become a properwayang orang. We need to rent unique costumes that are hard tofind and learn wayang orang makeup that can take a long time tomaster. Using image translation can make it easier for everyone to have the opportunity to see themselves as wayang orang.This study aims to translate human faces into wayang orang by adding makeup and accessories using the Generative Adversarial Network (GAN) and using unpaired dataset consist of 1216 data trains and 240 data tests. The challenge of this research is tomaintain the image background and the facial identity componentin the input image. This research uses quantitative testing employFID, KID, and IS to evaluate the quality of the generated image from the generator. The experimental result from this study is that UGATIT has the better result from DCLGAN based on the value from Inception Score, FID, and KID. Based on the IS, FID,and KID, the UGATIT has better results than the DCLGAN. Theresults from the UGATIT in IS, FID, and KID in the following order are 2.414, 0.810, and 6.221, which means UGATIT can perform better than the DCLGAN.
Keywords: image translation, GAN, unpaired dataset, wayang orang.

1. Pendahuluan

Latar Belakang

Kesenian wayang orang merupakan salah satu pertunjukan Indonesia yang identik dengan budaya tradisional Jawa khususnya Jawa Tengah. Namun, seiring perkembangan zaman kebudayaan Indonesia semakin luntur. Di era digitalisasi, budaya asing membuat budaya tradisional yang menjadi jati diri bangsa tergerus dan ditinggalkan oleh sebagian masyarakat, khususnya kalangan muda. Sedangkan hiburan tradisional harus dilestarikan untuk mengimbangi perkembangan zaman. Penelitian ini menjadi salah satu cara penulis untuk menjaga kelestarian warisan budaya indonesia di tengah modernisasi dengan menggabungkan teknologi dan budaya tradisional sehingga dapat menyesuaikan dengan perkembangan zaman.

Salah satu cara untuk menggabungkan teknologi dan budaya tradisional dengan cara image translation. image translation dapat dibangun menggunakan jaringan GAN. *Generative Adversarial Networks* (GAN) merupakan neural network yang digunakan untuk *unsupervised learning*. Ada banyak jenis implementasi GAN yang berbeda. Ada beberapa jenis pengaplikasian GAN antara lain pada *video prediction*, translasi tulisan menjadi citra, dan *image to image translation*.

Penelitian mengenai translasi citra wayang dan manusia belum ada sebelumnya, namun perkembangan penelitian yang serupa dengan topik penelitian ini yaitu penelitian mengenai image translation diawali oleh Leon A. Gatys pada tahun 2015, penelitian yang berfokus pada bidang kesenian [1], dan penelitian [2, 3, 4, 5] mampu melakukan tugas translasi antara citra manusia, boneka, dan anime. Penelitian ini mengadopsi jaringan GAN tepatnya arsitektur U-GAT-IT [3], dengan melakukan perubahan dengan menambah atribut aksesoris wayang dan gaya sambil mempertahankan atribut pada citra masukan seperti rambut, pose dan latar belakang. menggunakan dataset citra manusia dan citra wayang orang. Sehingga dengan menggunakan arsitektur U-GAT-IT diharapkan saat ada data baru pada sistem, sistem tetap bisa melakukan translasi data tersebut dengan baik.

Dengan adanya penelitian translasi citra berbasis GAN ini diharapkan dapat membantu para seniman, menjadi referensi baru pada atribut penokohan wayang. Bagi penikmat seni dan masyarakat awam tidak lagi hanya sebatas membayangkan wujud diri menjadi figur wayang melainkan dapat melihat langsung dari hasil translasi citra. Pengembangan riset ini menambah tugas baru yang dapat dilakukan dari model UGATIT yaitu translasi citra manusia dan wayang orang dan dapat memberikan dataset baru wayang orang untuk digunakan untuk keperluan riset lainnya.

Topik dan Batasannya

Berdasarkan dari latar belakang diatas, maka rumusan masalah pada penelitian ini adalah

1. Bagaimana membangun model yang dapat melakukan translasi antara citra wajah manusia dan wayang orang?
2. Bagaimana hasil gambar dalam translasi antara citra wajah manusia dan citra wayang orang?

Adapun batasan masalah dari penelitian ini sebagai berikut:

1. Dataset yang digunakan pada penelitian ini yaitu citra dengan format JPEG atau PNG.
2. Citra wajah dan wayang orang dengan *gender* pria.
3. Dataset citra wajah menggunakan dataset CelebA
4. Model translasi citra mengadopsi arsitektur UGATIT.
5. Citra masukan berbentuk persegi, berukuran 256x256 pikse.
6. Berfokus pada model translasi citra antara manusia dan citra wayang orang.

Tujuan

Berdasarkan topik dan batasan, tujuan dari penelitian ini antara lain :

1. Membangun model yang dapat melakukan translasi citra antara wajah manusia dan wayang orang.
2. Menganalisis hasil gambar dalam translasi antara citra wajah manusia dan citra wayang orang.

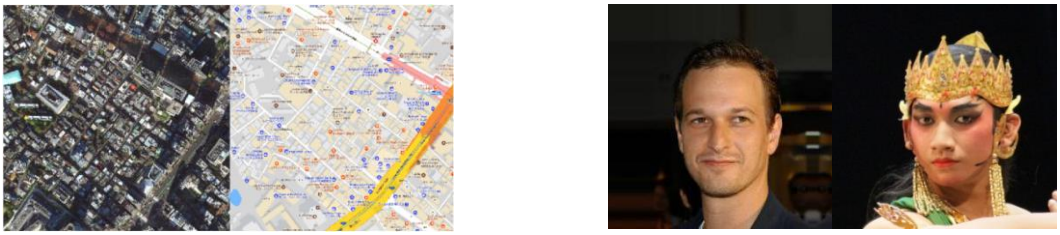
2. Studi Terkait

2.1 *Generative Advesarial Network*

Generative Adversarial Network (GAN) adalah arsitektur untuk melatih model generatif dengan memperlakukan masalah *unsupervised learning* yang menggunakan *supervised loss* sebagai bagian dari *training*. Jaringan GAN melakukan pembelajaran dengan memperoleh *backpropagation signals* melalui proses kompetitif yang melibatkan sepasang jaringan. Jaringan *generator* sebagai pembangkit citra sintesis sementara jaringan *discriminator* sebagai jaringan yang bertugas untuk membedakan antara citra asli dan citra yang dihasilkan oleh *generator*. Jika GAN berhasil dilatih, maka *generator* dapat menghasilkan citra yang sulit dibedakan dari citra wayang orang asli sehingga membingungkan jaringan *discriminator*.

2.2 *Image to Image Translation*

Image to image translation bertujuan untuk mengubah citra dari satu domain dengan mengubah gaya atau karakteristik dari citra ke domain lain. *Image to image translation* dapat dipelajari dengan *supervised* dan *unsupervised*. Pada penelitian ini menerapkan *unsupervised learning*, karena dataset citra merupakan citra yang independen atau tidak saling berkaitan (*unpaired*). Tujuan metode *unsupervised learning* adalah untuk menerjemahkan antara domain yang berbeda dengan menggunakan citra yang tidak berlabel tanpa menetapkan kaitan antar citra, sehingga dapat menghemat *cost* dari data berlabel. Pendekatan yang menggunakan *unsupervised learning* yaitu UNIT [6], CycleGAN [5] dan DualGAN [7].



A. Paired *B. Unpaired*

Gambar 1. Contoh Dataset Paired dan Unpaired

Dalam penerapannya pada penelitian [5], begitu banyak yang dapat diaplikasikan dari *image to image translation*, seperti *collection style transfer,object transguration, season transfer* dan *photo enhancement*. Pada tahun 2016 translasi citra berbasis GAN pertama kali diperkenalkan oleh Isola et al [16], ketika citra yang dihasilkan oleh generator diadaptasi dari citra masukan, proses tersebut dinamakan translasi citra. Arsitektur CycleGAN [5], memiliki dua "*Cycle Consistancy Losses*" yang mampu melakukan translasi dari satu domain menjadi domain target dan jika dibalik, akan kembali seperti semula. Pada penilitian *image to image translation* dengan translasi kuda menjadi zebra dengan metode CycleGAN mencapai skor FID senilai 89,7. sedangkan jika menggunakan metode UGATIT [3] menggunakan metriks evaluasi *Kernel Inception Distance* dengan skor KID mencapai 7,06. Pada translasi swafoto menjadi anime dengan metode UGATIT mencapai skor KID 11,67. Dan pada tranlasi anjing menjadi kucing mencapai skor KID yaitu 7.07 menunjukkan bahwa model UGATIT mampu melakukan berbagai macam *task image-to-image translation*.

2.2 *Class Activation Map* (CAM)

Zhou et al. memperkenalkan CAM [8] menggunakan *global average pooling* pada CNN. CAM pada kelas kelas tertentu menunjukkan daerah diskriminatif citra oleh CNN untuk menentukan kelas. Pada penelitian ini, model diarahkan untuk perubahan daerah diskriminatif citra secara intensif dengan membedakan dua domain menggunakan pendekatan CAM. Dan juga digunakan penggabungan *max pooling* untuk membuat hasil menjadi lebih baik.

2.3 *Inception Score* (IS)

Inceptions Score pertama kali diperkenalkan oleh Salimans et al [10]. IS menjadi evaluasi alternatif dari evaluasi dengan *annotator* manusia dalam kualitas citra, khususnya citra keluaran dari model GAN. Nilai dari inception skor berdasarkan citra yang beragam (setiap gambar wayang orang berbeda satu dengan lainnya),dan setiap citra hasil terlihat gambar target yang realistis. Batasan nilai dari IS pada rentang [1, ∞]. Semakin besar nilai IS, menunjukkan kualitas gambar yang semakin baik.

$$IS = exp(E_x\ D_{KL}(p(y|x)||p(y)))$$

2.4 *Fré chet Inception Distance* (FID)

FID bertujuan untuk mengevaluasi kualitas citra hasil dari generator dan mengevaluasi performansi GAN. Distribusi data dimodelkan menggunakan distribusi Gaussian multivariat dengan mean μ dan kovarians $\sum$. Menurut Martin Heusel et al [9] evaluasi menggunakan FID lebih baik dibandingkan metode evaluasi yang telah ada sebelumnya. Karena FID mengevaluasi kumpulan data citra sintesis dibandingkan dengan kumpulan citra asli dari domain target. Skor FID antara citra asli (r) dan citra hasil (g) dihitung dengan formula berikut:

$$FID(r,g) = ||\mu_r - \mu_g|| + Tr(\sum_r + \sum_g - 2(\sum_r \sum_g)^{\frac{1}{2}}$$

Keterangan:
- r = Gambar asli (*real*)
- g = Gambar hasil (*generated*)
- μ = Mean dari fitur citra
- $\sum$ = Matriks kovarians dari fitur citra

Skor FID yang lebih rendah menunjukkan gambar berkualitas lebih baik, berdasarkan dari citra yang dihasilkan (sintetis) lebih mirip dengan citra asli. Nilai terendah dari Skor FID yaitu bernilai 0 jika gambar yang dibandingkan adalah gambar yang identik. Cara kerja dari FID yaitu dengan mengukur jarak antar *activation distributions*.

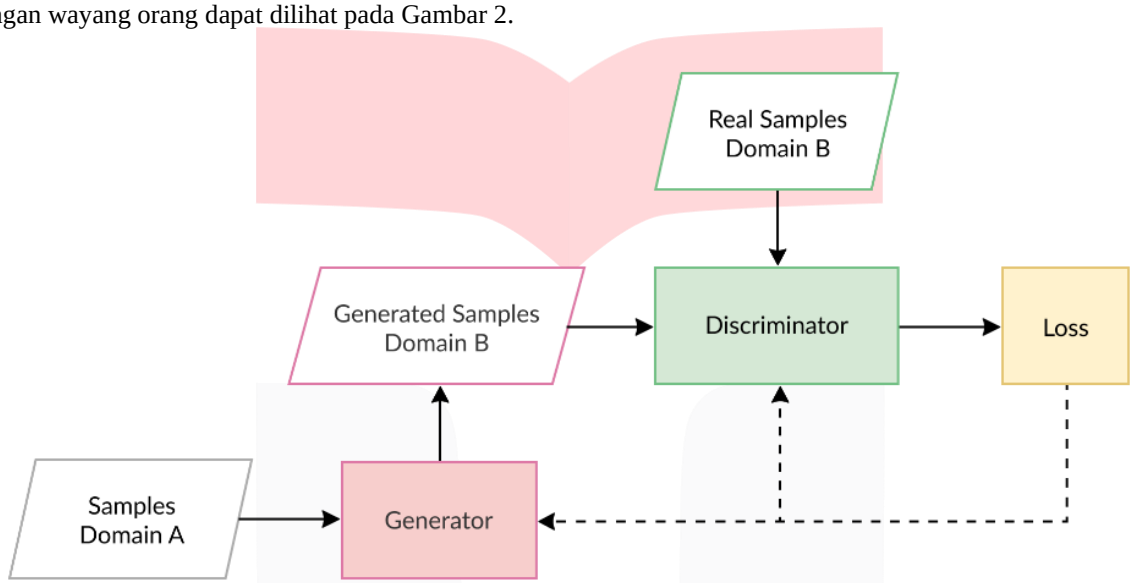
2.5 *Kernel Inception Distance* (KID)

Evaluasi *Kernel Inception Distance* [11] mirip dengan FID. FID dan KID menggunakan varian uji dua sampel dalam *feature space* "perseptual" yang dipelajari, *Inception pool3 space*, untuk menilai kecocokan distribusi. Perbedaannya, KID menghitung kuadrat *Maximum Mean Discrepancy* (MMD) antara representasi *inception*. Terlebih lagi, kelebihan dari KID yaitu merupakan *unbiased estimator*.

3. Sistem yang Dibangun

3.1 Desain Sistem

Pada penelitian ini mengadopsi arsitektur UGATIT [3]. Arsitektur UGATIT memiliki sepasang generator dan sepasang discriminator, yang bertugas melakukan translasi citra dari domain asal ke domain target kemudian dikembalikan kembali ke domain asal. Namun pada penelitian ini hanya akan membahas perubahan satu domain ke domain target saja. Rancangan sistem yang digunakan untuk mentranslasikan antara gambar wajah manusia dengan wayang orang dapat dilihat pada Gambar 2.



Gambar 2. Desain Sistem Proses Pelatihan GAN

Alur dari desain sistem diawali dengan memasukan citra domain asal (wajah orang) ke jaringan *generator*. Pada jaringan *generator* bertugas untuk menghasilkan citra dengan domain target yaitu citra domain wayang orang. Setelah itu, pada jaringan *discriminator* akan dibandingkan citra palsu *generator* dan citra asli wayang orang. Kemudian dilakukan perhitungan fungsi *loss* yang kemudian dilakukan pembaruan ke jaringan *generator* dan juga *discriminator*.

Tabel 1. Hyperparameter Model

Optimizer	*Learning Rate*	Beta1	Beta2	*Adversarial* GAN Type	Bobot *Identity*	Bobot *Cycle*	Bobot CAM	*Batch Size*
Adam	0,0001	0,5	0,999	LSGAN	10	10	1000	1

Penggunaan *Adaptive moment estimation* (Adam) berperan untuk mengoptimalkan model dengan memperbarui bobot dari jaringan pada proses pelatihan. Adam relatif mudah untuk dikonfigurasi dimana parameter konfigurasi bekerja dengan baik pada sebagian besar masalah. Konfigurasi *hyperparameter* model yang digunakan berdasarkan rekomendasi dari penelitian [3, 5] meliputi *adam optimizer, learning rate, beta1, beta2, identity weight, cycle weight, cam weight* dan tipe *adversarial* GAN dengan rincian pada table 1.

3.2 Dataset

Penelitian ini menggunakan *dataset* yang tidak saling berkaitan (*unpaired dataset*) yaitu citra wajah (domain asal) dan citra wayang orang (domain target), karena sulitnya mendapatkan citra yang berpasangan (*paired dataset*). Kumpulan gambar wayang orang pada penelitian ini didapatkan dari komunitas wayang yang bernama Wayang Kautaman. Ditambah lagi dengan dengan metode *scraping image*. Sedangkan citra wajah menggunakan dataset CelebA [12] digabungkan dengan wajah orang Indonesia. Penggabungan citra ini agar dataset citra bervariasi dan tidak bias. Citra yang digunakan pada penelitian ini dibatasi hanya *gender* pria saja. Sampel citra dapat dilihat pada gambar dibawah ini. Semakin banyak jumlah citra dan semakin bervariasi gambar dari citra akan berpengaruh baik pada model.

Gambar 3. Sampel Citra CelebA (Domain A)

Gambar 4. Sampel citra wajah Indonesia (Domain A)

Gambar 5. Sampel Citra Wayang Orang (Domain B)

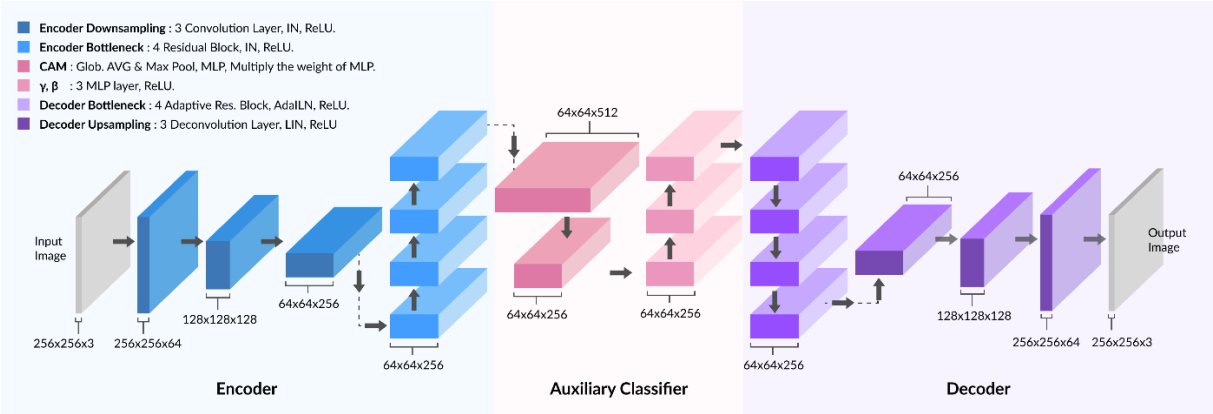
Setelah *Dataset* terkumpul, kemudian dataset dibagi menjadi dua bagian yaitu data latih dan data uji. Pembagian *dataset* citra wajah pada data latih, dan data uji menggunakan prinsip pareto dengan rasio 80:20. Maka jumlah data latih dan data uji yaitu sebanyak 976, dan 240 citra. Begitupun *dataset* pada citra wayang orang, data latih dan data uji masing-masing berjumlah 240 citra.

Sebelum tahap pelatihan dilakukan, penting untuk melakukan *facial alignment* pada semua citra dengan merotasi gambar berdasarkan *keypoints* dari wajah. Pada penelitian ini proses *facial alignment* dilakukan secara manual dengan menyamakan titik pada bagian mata dan mulut. *Facial alignment* disini bertujuan untuk menghindari perubahan bentuk wajah citra asal dengan mempertahankan struktur wajah citra asal pada proses pelatihan nantinya. Proses terakhir dari sebelum proses pelatihan yaitu *dataset* citra kemudian dipotong berbentuk persegi dan langkah terakhir ukuran citra diubah menjadi 256×256 piksel (*resize*).

3.3 *Generator*

Generator dari arsitektur UGATIT dapat dilihat pada Gambar 6. Terbagi oleh *encoder*, *auxiliary classifier*, dan *decoder*. *Encoder downsampling* memiliki dua layer konvolusi dengan *strides* berukuran dua, *padding* berukuran satu, sedangkan pada *encoder bottleneck* terdiri dari empat *residual block*. Bagian *Auxiliary classifier* dilatih untuk mempelajari bobot dari *feature map* pada domain asal menggunakan *global average pooling* dan *global max pooling*. bagian *decoder bottleneck* terdapat empat *residual block* dan bagian terakhir yaitu *decoder up-sampling* terdiri dari tiga layer konvolusi. Pada *decoder upsampling* terakhir terdapat satu layer konvolusi dengan fungsi aktivasi Tanh. Penggunaan Tanh untuk menyamakan *range* dari tiap piksel citra [-1, 1].

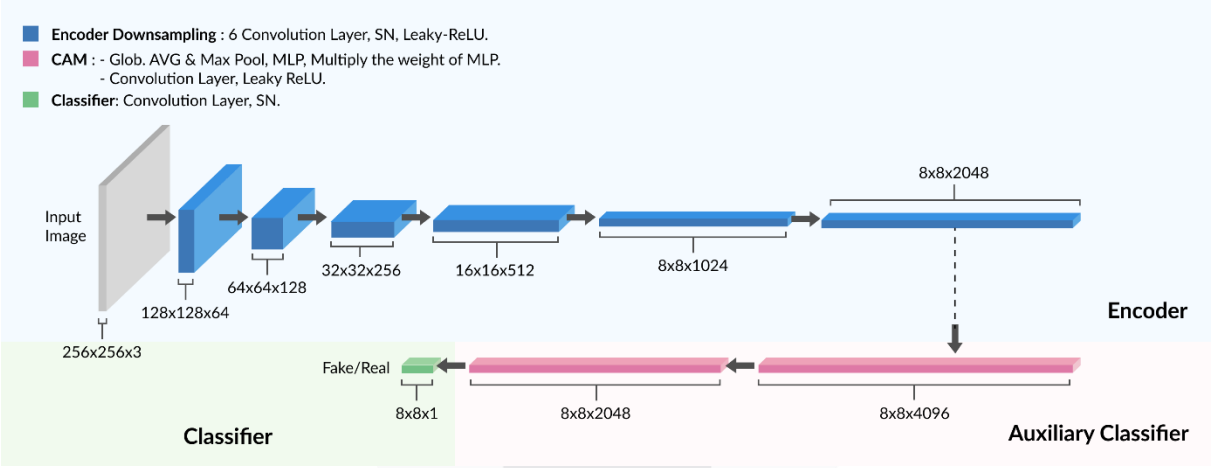
Normalisasi yang diterapkan pada bagian *encoder* yaitu menggunakan *Instance Normalization* (IN), dan AdaLIN [3] pada *decoder*. Adapula fungsi aktivasi yang diterpakan pada jaringan *generator* ini yaitu ReLU dan tanh. Menurut junho kim et. al [3], fungsi normalisasi AdaLIN ini menggabungkan kelebihan dari *Adaptive Instance Normalization* [17] dan *Layer Normalization* (LN) [16]. *Adaptive Instance Normalization* sendiri memiliki kemampuan untuk mengubah dan menyimpan konten gambar secara selektif.



Gambar 6. Arsitektur *Generator* GAN

Mula-mula, citra domain wajah orang akan melewati bagian *encoder* yang bertugas untuk mengekstraksi fitur dan menghasilkan *feature map*. Selain itu, downsampling dari dimensi spasial membuat pengumpulan informasi, dari area yang lebih luas, menjadu lebih efisien. Terdapat blok residual pada *encoder bottleneck* yang bertugas untuk membantu model dalam mempelajari transformasi efektif dari domain asal (wajah) ke domain target (wayang orang). Kemudian pada bagian *auxiliary classifier* terdapat *attention map* yang berguna untuk membantu *generator* untuk fokus pada bagian yang membedakan dari domain wajah dan domain wayang orang. Sehingga, membantu model dalam mengubah bagian tertentu saja. Selanjutnya, pada bagian *decoder* akan membuat citra domain target (wayang orang) dari *embedded feature*. Keluaran dari jaringan generator ini yaitu citra palsu berdomain wayang orang.

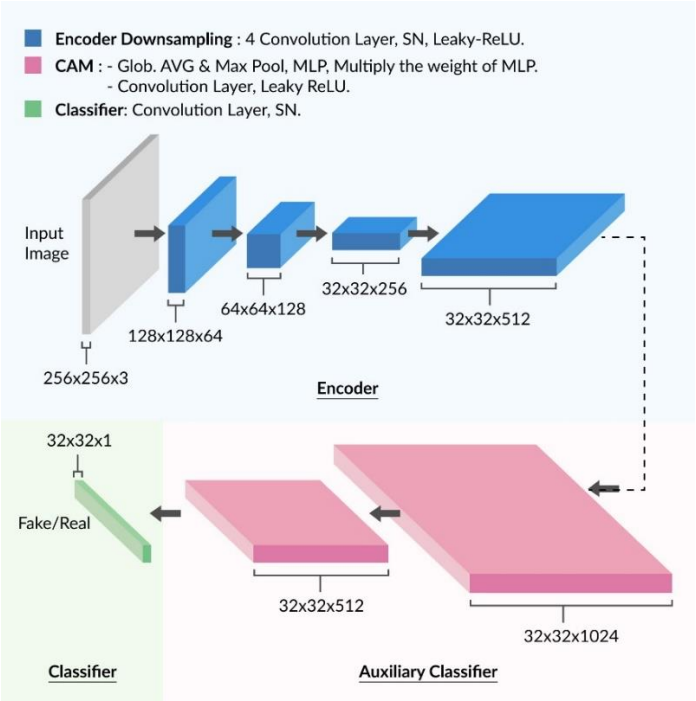
3.4 *Discriminator*



Gambar 7. Arsitektur *Discriminator Global* GAN

Berbeda dengan model *generator*, model ini terbagi oleh *encoder*, *auxiliary classifier*, dam *classifier*. Pada masing masing *encoder* menggunakan *instance normalization*, dan *activation function* Leaky ReLU. Model *discriminator* dan *auxiliary classifier* dilatih untuk membedakan apakah citra inputan adalah citra sintetik atau citra target. Pada *discriminator* menggunakan patch-GAN[8] yang mengklasifikasi citra sintetik atau asli dengan ukuran 70x70 (lokal) dan 256x256 (global). Alur dari jaringan *discriminator* diawali dengan masukan citra sampel, dan keluaran dari jaringan *discriminator* apakah gambar masukan adalah gambar asli atau gambar palsu. Jaringan *discriminator* memanfaatkan *attention feature maps* yang menggunakan bobot dari *encoded feature maps* yang telah dilatih oleh *auxiliary classifier*. Menggunakan *Spectral Normalization* (SN) [14] sebagai fungsi normalisasi. Pada *discriminator global* pada *encoder downsampling* berjumlah lebih banyak dari *discriminator* lokal, sebanyak enam layer sedangkan *discriminator* lokal hanya empat layer.

Attention module diaplikasikan pada jaringan *discriminator* dan *generator*. Sehingga *attention module* pada jaringan *discriminator* mengarahkan jaringan *generator* untuk fokus pada bagian penting dari citra realistis. Dan juga *attention module* pada jaringan *generator* memberikan perhatian lebih (fokus) pada bagian yang membedakan dari domain lain.



Gambar 8. Arsitektur Discriminator Lokal GAN.

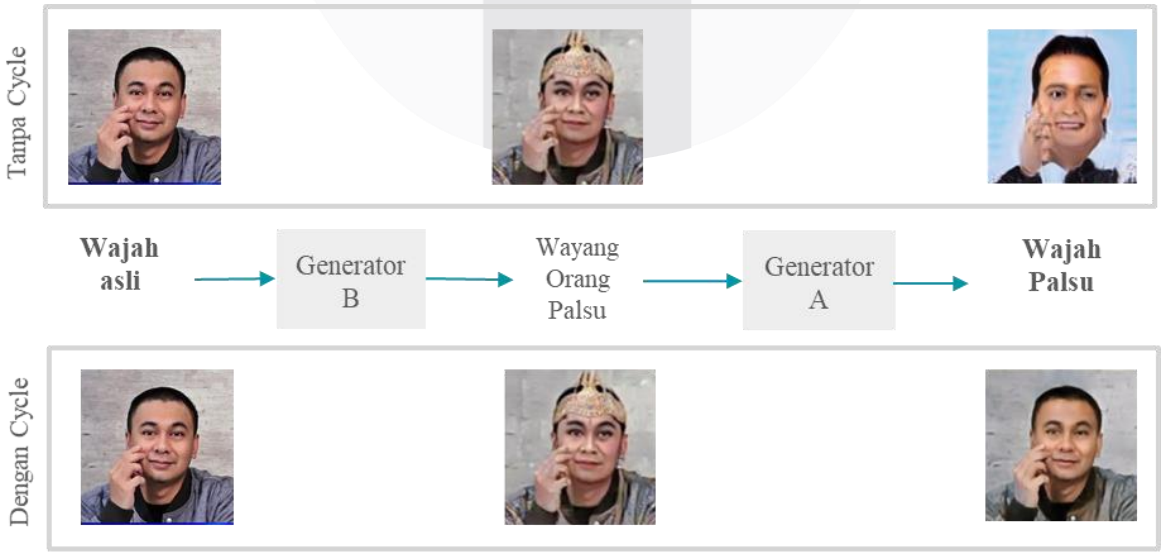
3.5 Fungsi Loss

Untuk membedakan antara gambar palsu dan gambar asli, model melakukan pembelajaran dengan menggunakan fungsi *adversarial loss*. Berdasarkan penelitian [13] *adversarial loss* LSGAN mampu menstabilkan model dibandingkan tipe adversarial regular GAN dengan. Namun, masalah GAN yang seringkali terjadi yaitu masalah *mode collapse* yaitu saat model hanya menghasilkan gambar yang sama terus menerus. Maka dari itu, tidak hanya menggunakan *fungsi loss adversarial* pada penelitian ini menggunakan tiga fungsi loss tambahan yaitu cycle loss, CAM loss, dan identity loss.

Adversarial loss, tepatnya bertipe LSGAN digunakan untuk menyamakan distribusi antara citra wayang asli (target) dan citra wayang hasil dari *generator* (sintesis).

$$\mathcal{L}_{GAN} = (E_{x\to Xt}\,[log(Dt(X))] + Ex \sim Xs\,[log(1 - Dt(G_{s\to t}(X))])$$

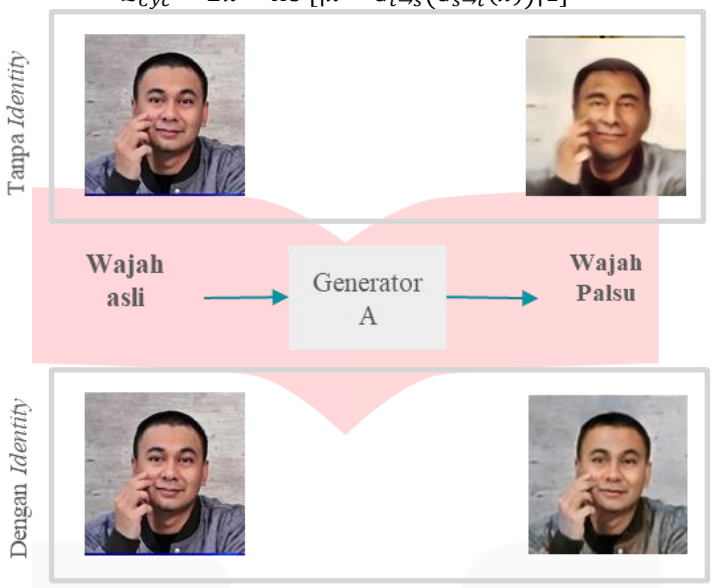
$$\mathcal{L}_{LSGAN} = (E_{x\to Xt}\,[(D_t(X))^2] + E_{x\to Xs}[(1 - D_t(G_{s\to t}(X)))^2])$$



Gambar 9. *Cycle loss*.

Cycle loss, untuk mengatasi masalah yang seringkali terjadi pada GAN yaitu *mode collapse* dengan menggunakan *cycle consistency constraint* pada *generator*. Pada Gambar 9, Contoh jika model tidak menggunakan cycle loss pada bagian atas kanan, sedangkan pada bagian bawah jika model menggunakan cycle loss. Alurnya diawali dengan diberikan citra wajah asli sebagai *input* citra dari generator B menghasilkan citra palsu wayang orang. Kemudian akan ditranslasi kembali ke domain asal (wajah biasa), dengan cara menjadikan citra keluaran *generator* B sebagai citra masukkan dari *generator* A. Setelah itu wajah palsu akan dibandingkan dengan wajah wajah asli, proses ini untuk memastikan citra palsu berdomain wayang dapat ditranslasi kembali mendekati citra asal.

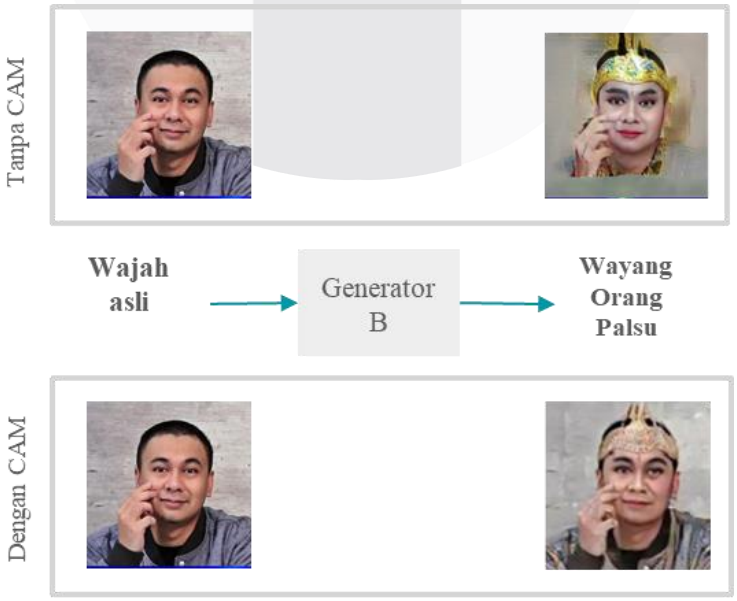
$$\mathcal{L}_{Cyc} = Ex \sim Xs\ [|x - G_{t\to s}(G_{s\to t}(x))|1]$$



Gambar 10. *Identity Loss*

Identity loss, untuk memastikan distribusi warna dari citra masukan dan citra keluaran serupa, dengan membatasi konsistensi identitas pada *generator*. Diberikan sampel dari citra target, setelah translasi secara terbalik, citra awal tidak mengalami perubahan. Penggunaan *identity loss* yaitu dengan memasukan citra berdomain wajah ke generator dengan domain yang sama, bertujuan untuk *identity mapping*. Pada Gambar 10, Contoh jika model tidak menggunakan *identity loss* pada bagian kanan atas, sedangkan pada bagian bawah jika model menggunakan CAM loss.

$$\mathcal{L}_{Identity}^{s\to t} = E_{x\to Xt}\ [|x - G_{s\to t}(x)|1]$$



Gambar 11. *CAM loss*

CAM *loss*, dengan memanfaatkan informasi dari *auxiliary classifer* pada jaringan *generator* dan *discriminator*. Diberikan sampel citra dari domain manusia dan wayang. *Generator* dan *discriminator* perlu mengetahui pada bagian gambar yang perlu ditingkatkan, dan apa yang membuat perbedaan antara dua domain pada *current state*. Pada Gambar 11, Contoh jika model tidak menggunakan CAM loss pada bagian kanan, sedangkan pada bagian bawah jika model menggunakan CAM loss.

$$\mathcal{L}_{CAM}^{s\to t} = -(E_{x\to Xs}[\log(\eta_s(x))] + E_{x\to Xt}[\log(1-\eta_s(x))])$$

$$\mathcal{L}_{CAM}^{Dt} = E_{x\to Xt}[\eta_{Dt}(x)]^2 + E_{x\to Xs}[(1-\eta_{Dt}(G_{s\to t}(x)))^2]$$

Dengan menggabungkan *encoders, decoders, discriminators* dan *auxiliary classifer*. Maka total *loss* didefinisikan sebagai berikut:

$$\mathcal{L}_{LSGAN} = \mathcal{L}_{LSGAN}^{s\to t} + \mathcal{L}_{LSGAN}^{t\to s}$$

$$\min_{G_{s\to t},G_{t\to s},\eta s,\eta t} \max_{Ds,Dt,\eta Ds,\eta Dt} = \lambda_1\mathcal{L}_{LSGAN} + \lambda_2\mathcal{L}_{Cycle} + \lambda_3\mathcal{L}_{Identity} + \lambda_4\mathcal{L}_{CAM}$$

Keterangan:

x	= sampel citra
D	= *discriminator*
G	= *generator*
$s \to t$	= domain *source* ke *target*
$t \to s$	= domain *target* ke *source*
η_t	= *auxiliary classifier* pada *generator*
η_s	= *auxiliary classifier* pada *discriminator*

Dimana
$\lambda_1 = 1.$
$\lambda_2 = 10.$
$\lambda_3 = 10.$
$\lambda_4 = 1000.$

4. Evaluasi

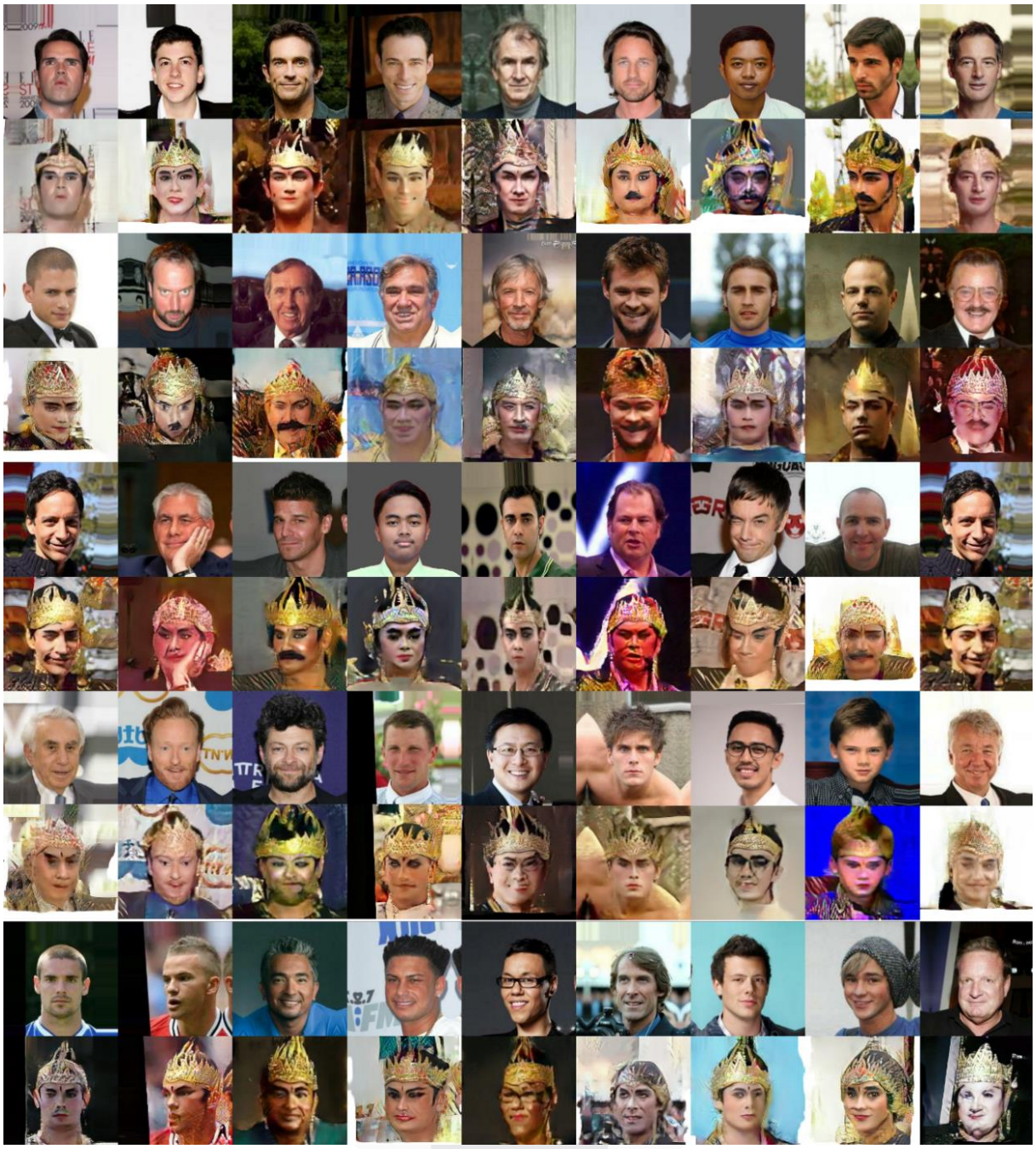
Pengujian dilakukan untuk mengetahui kinerja dari model. Faktor penting dalam mengevaluasi kualitas gambar salah satunya berdasarkan seberapa mirip dan realistis gambar yang dihasilkan. Tetapi, gambar buram pun masih bisa terlihat realistis, maka ketajaman gambar juga menjadi faktor yang perlu diperhatikan. Faktor lainnya adalah *generator* yang mampu menghasilkan gambar-gambar yang bervariasi, bukan hanya menghasilkan gambar dalam jumlah banyak namun dengan gambar yang sama. Tantangan dalam translasi citra wajah menjadi wayang orang adalah ketika harus mengubah gaya dari suatu citra selagi mempertahankan identitas wajah dan latar belakang dari gambar domain asal. Evaluasi gan tidak cukup dengan melihat fungsi *loss*.

Pada penelitian ini untuk mengukur performa model dan gambar yang dihasilkan, digunakan metode FID, IS, dan KID sebagai evaluasi kuantitatif menggunakan data uji sebanyak 240 citra wajah. Citra yang menjadi data uji adalah data baru dan belum pernah digunakan pada proses pelatihan.

Model UGATIT bandingkan dengan model DCLGAN, karena memiliki kemampuan yang sama yaitu *image-to-image translation* dengan pembelajaran tanpa pengawasan. Perlakuan pada model DCLGAN sama dengan model UGATIT dilakukan *training* sebanyak 500,000 iterasi dan konfigurasi *hyperparameter* model yang sama. Model DCLGAN mampu melakukan translasi citra pada berbagai macam objek, seperti translasi antara kuda dan zebra, apel dan jeruk, foto dan lukisan.

4.1. Hasil Pelatihan

Gambar 12 merupakan sampel gambar hasil pelatihan model UGATIT sebanyak 50 *epoch* atau 500,000 iterasi. Tujuan dari model UGATIT yaitu mentranslasikan gambar *foreground* ke domain wayang orang, sementara gambar *background* tetap pada domain asal atau tidak mengalami perubahan. Citra sintesis yang diharapkan



Gambar 12. Sampel Gambar Hasil Pelatihan model UGATIT

Gambar diatas merupakan contoh pasangan dari gambar asal dan gambar palsu. Gambar asal berada pada baris pertama dan gambar palsu di baris kedua begitu pula pada pasangan baris selanjutnya. Dari keseluruhan hasil, terdapat beberapa citra wajah yang awalnya senyum dengan bibir terbuka setelah ditranslasasi menjadi senyum tertutup karena data latih hanya memiliki sedikit citra wayang orang dengan senyum terbuka. Untuk mengatasi hal tersebut bisa dengan memperbanyak data latih dengan citra wajah tersenyum. Pada baris terakhir kedua dari kanan, citra domain orang menggunakan topi, model tetap mampu membentuk mahkota.

4.2 Hasil dan Analisis Gambar Hasil Uji

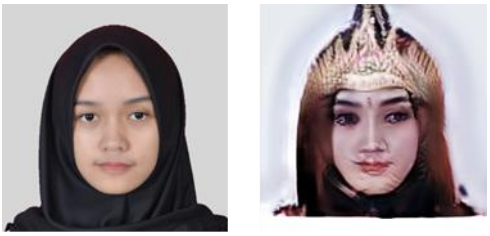
Masalah dengan model generatif adalah bahwa tidak ada cara objektif untuk mengevaluasi kualitas gambar yang dihasilkan. Umumnya dilakukan evaluasi subjektif manusia secara berkala untuk menentukan akhir dari model. Gambar dibawah ini merupakan gambar hasil uji dari sampel gambar yang dengan kondisi yang bervariasi, seperti memakai kacamata, memakai masker, dan lebih dari satu orang dalam satu gambar, kondisi gambar yang beragam ini untuk mengevaluasi kualitas dari gambar dari tiap model.

Input Image DCLGAN UGATIT

Gambar 13. Sampel Gambar Hasil Uji.

Pada gambar 13, menunjukkan bahwa gambar hasil uji dari model DCLGAN terlihat mahkota pada kepala setiap orang, namun model DCLGAN mengubah keseluruhan warna pada gambar menjadi berwarna kekuningan, Tidak hanya warna, gambar hasil uji dari model DCLGAN terdapat gangguan garis pada sisi bagian bawah gambar. Selain itu, latar belakang pada gambar juga telah berubah dari gambar masukkan.

Gambar dengan hasil paling baik dari model UGATIT dapat dilihat pada contoh Gambar 13 di baris pertama, gambar dapat terbentuk mahkota dengan jelas dan juga terbentuk anting. Begitu juga dengan riasan wajah dapat ditransfer dengan baik, lipstik warna merah pada bagian bibir, alis terlihat lebih gelap, cambang dan juga tanda pada dahi. Model UGATIT dapat mentranslasi gambar tanpa mengubah konten (identitas dan latar belakang) dari gambar masukkan. Pada sampel gambar hasil uji baris ke empat, gambar orang memakai masker tidak terdapat perona bibir karena bibir yang tertutup masker. Pada baris terakhir model UGATIT juga mampu melakukan translasi pada lebih dari satu orang dalam satu gambar.



Gambar 14. Sampel Citra Hasil Pada Wanita Menggunakan Hijab

4.2. Hasil dan Analisis berdasarkan Hasil Kuantitatif

Untuk mengevaluasi hasil citra uji dibuktikan dengan melakukan evaluasi kuantitatif pada model UGATIT dan juga model DCLGAN. Evaluasi kuantitatif yang diterapkan pada penelitian ini yaitu IS, FID dan KID. Cara kerja dari IS adalah dengan mengukur kumpulan citra palsu hasil keluaran *generator*, bertujuan untuk memastikan kumpulan citra bervariasi atau unik. Selain itu, untuk memastikan kualitas dari citra

. Tabel 1. Hasil Uji Kuantitatif

Model	*Inception Score* (IS) ↑	FID x 100 ↓	KID x 100 ↓
	500,000 iterasi		
A. UGATIT	**2.414**	**0.924**	**4.357**
B. DCLGAN	1.670	2.189	22.13

Berdasarkan hasil eksperimen pada tabel 2. Model UGATIT memberikan hasil yang lebih baik ditunjukkan dengan Skor IS yang lebih tinggi dan skor FID dan KID yang lebih rendah dibandingkan DCLGAN. Dengan nilai IS 2.414, FID 0.924 dan KID 4.357 pada model UGATIT dan IS 1.670, FID 2.189 dan KID 22.13 pada DCLGAN.

5. Kesimpulan dan Saran

Proses dari sistem ini diawali dengan pengumpulan dataset. Kemudian pada proses *preprocessing*, ada baiknya untuk melakukan *facial alignment* sebelum proses pelatihan. Berdasarkan hasil dan analisis dari eksperimen yang telah dilakukan, dapat disimpukan bahwa dari hasil pelatihan, gambar terbaik adalah gambar dengan riasan wajah dan kostum yang berhasil ditambahkan ke wajah. Selain itu, gambar hasil yang mempertahankan latar belakang dan juga identitas wajah dari gambar asal. Pada bagian hasil dan analisis dari gambar hasil uji model mampu melakukan translasi citra pada berbagai macam kondisi gambar, seperti orang memakai kacamata, masker, dan lebih dari satu orang dalam satu gambar. Model UGATIT mampu menghasilkan kualitas gambar lebih baik dan lebih menyerupai wayang orang dibandingkan DCLGAN. Dibuktikan dengan evaluasi menggunakan *Inception Score* (IS). Nilai IS dari model UGATIT yaitu 2.414, lebih tinggi dibandingkan DCLGAN. Selain itu, skor FID dan KID dari model UGATIT yaitu 0.924 dan 4.357, sedangkan skor FID dan KID senilai 2.189 dan 22.13 dari model DCLGAN. Dari hasil evaluasi menunjukkan UGATIT model dapat bekerja dengan baik dalam mentranslasikan wajah orang menjadi wayang orang.

Penelitian ini dapat menjadi *novel task* dari UGATIT model, translasi citra manusia menjadi wayang orang. Dataset wayang orang pada penelitian ini dapat digunakan untuk kebutuhan riset dimasa mendatang, dan untuk peneliti selanjutnya juga dapat memperbanyak jumlah citra wayang orang dari dataset ini. Pengembangan penelitian selanjutnya dapat dilakukan menggunakan dataset citra dengan *gender* wanita.

REFERENSI

[1] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. A neural algorithm of artistic style. CoRR, abs/1508.06576, 2015.

[2] Aaron Gokaslan, Vivek Ramanujan, Daniel Ritchie, Kwang In Kim, and James Tompkin. Improving shape deformation in tanpa pengawasan image-to image translation. In Proceedings of the European Conference on Computer Vision (ECCV), September 2018.

[3] Junho Kim, Minjae Kim, Hyeonwoo Kang, and Kwanghee Lee. UGAT-IT: unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation. CoRR, abs/1907.10830, 2019.

[4] Taeksoo Kim, Moonsu Cha, Hyunsoo Kim, Jung Kwon Lee, and Jiwon Kim. Learning to discover cross-domain relations with generative adversarial networks, 2017.

[5] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. InProceedings of the IEEE International Conference on Computer Vision (ICCV), Oct 2017.

[6] Ming-Yu Liu, Thomas Breuel, and Jan Kautz. Unsupervised image-toimage translation networks. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, Advances in Neural Information Processing Systems, volume 30, pages 700{708.Curran Associates, Inc., 2017.

[7] Zili Yi, Hao Zhang, Ping Tan, and Minglun Gong. Dualgan: Unsupervised dual learning for image-to-image translation. CoRR, abs/1704.02510,2017.

[8] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Imageto-image translation with conditional adversarial networks. CoRR, abs/1611.07004, 2016.

[9] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, Gunter Klambauer, and Sepp Hochreiter. Gans trained by a two timescale update rule converge to a nash equilibrium. CoRR, abs/1706.08500, 2017.

[10] Tim Salimans, Ian J. Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans. CoRR, abs/1606.03498, 2016.

[11] Mikołaj Bi´nkowski, Dougal J. Sutherland, Michael Arbel, and Arthur Gretton. Demystifying MMD GANs. In International Conference on Learning Representations, 2018.

[12] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In Proceedings of International Conference on Computer Vision (ICCV), December 2015.

[13] Xudong Mao, Qing Li, Haoran Xie, Raymond Y. K. Lau, Zhen Wang, and Stephen Paul Smolley. Least squares generative adversarial networks, 2017.

[14] Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. Spectral normalization for generative adversarial networks. In International Conference on Learning Representations, 2018.

[15] Junlin Han, Mehrdad Shoeiby, Lars Petersson, and Mohammad Ali Ar-min. Dual contrastive learning for tanpa pengawasan image-to-image transla-tion, 2021.

[16] Jimmy Lei Ba, Jamie Ryan Kiros, and Geo rey E. Hinton. Layer normalization, 2016.

[17] Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization, 2017.