

ANALISIS KINERJA PENYELARASAN WAJAH PADA PROSES REKONSTRUKSI WAJAH 3-DIMENSI (3D) DENGAN METODE MODIFIKASI POSITION MAP REGRESSION NETWORK

PERFORMANCE ANALYSIS OF FACE ALIGNMENT ON 3-DIMENSIONAL (3D) FACE RECONSTRUCTION USING MODIFIED POSITION MAP REGRESSION NETWORK)

Raka Cendekia Vandara¹, Suryo Adhi Wibowo², Koredianto Usman³

^{1,2,3} Universitas Telkom, Bandung

¹rakacendekia@student.telkomuniversity.ac.id, ²suryoadhiwibowo@telkomuniversity.ac.id,

³korediantousman@telkomuniversity.ac.id

Abstrak

Teknologi 3D *face reconstruction* tidak lepas dari perkembangan *face alignment*. Salah satu metode yang diterapkan dalam pembuatan teknologi *face alignment* adalah *Position Map Regression Network* (PRNet). Pada pengujian dengan dataset uji AFLW2000-3D, metode PRNet menjadi metode terbaik pada teknologi 3D *face reconstruction*, namun performansi *face alignment* pada metode ini tidak sebanding dengan beberapa metode baru. Pada Tugas Akhir ini dilakukan modifikasi pada konfigurasi arsitektur dan parameter PRNet. PRNet bekerja dengan mengandalkan representasi citra 2D pada ruang koordinat UV yang disebut *UV position map* yang merekam bentuk 3D dari sebuah wajah, sehingga diperoleh informasi geometri sebuah wajah hanya dari satu citra. Hasil keluaran sistem berupa penyelarasan titik - titik wajah pada 68 *landmark* wajah dan 45 ribu *point* wajah. Skema yang digunakan pada proses pengujian adalah skema *padding*, modifikasi, pengelompokan berdasarkan pose, dan pengelompokan berdasarkan *yaw angle*. Konfigurasi terbaik pada NME 2D didapatkan pada konfigurasi Modifikasi dengan nilai terbaik pada 68 *landmark* dan 45 ribu *point* masing - masing sebesar 3,63% dan 3,4%. Di sisi lain, performansi terbaik pada NME 3D didapatkan pada konfigurasi asli PRNet dengan nilai terbaik pada 68 *landmark* dan 45 ribu *point* masing - masing sebesar 5,08% dan 4,65%.

Kata kunci : PRNet, 3D *face reconstruction*, *face alignment*, *UV position map*.

Abstract

3D *face reconstruction technology* cannot be separated from the development of *face alignment*. One of the methods applied in making *face alignment technology* is the *Position Map Regression Network* (PRNet). In testing with the AFLW2000-3D test dataset, the PRNet method is the best method for 3D *face reconstruction technology*, but the performance of *face alignment* in this method is not comparable to some of the new methods. In this final project, modifications to the architecture configuration and parameters of PRNet are made. PRNet works by relying on a 2D image representation in the UV coordinate space called a *UV position map* which records the 3D shape of a face, so that geometric information of a face is obtained from only one image. The output of the system is the alignment of facial points on 68 facial landmarks and 45k thousand facial points. The scheme used in the testing process is the padding scheme, modification, grouping based on poses, and grouping based on *yaw angle*. The best configuration in NME 2D is obtained in the Modified configuration with the best value at 68 landmarks and 45 thousand points - 3.63% and 3.4%, respectively. On the other hand, the best performance on NME 3D is obtained in the original PRNet configuration with the best scores at 68 landmarks and 45 thousand points of 5.08% and 4.65%, respectively.

Keywords: PRNet, 3D *face reconstruction*, *face alignment*, *UV position map*.

1. Pendahuluan

Perkembangan teknologi dan implementasi 3D terus ditingkatkan seiring dengan berkembangnya teknologi khususnya di bidang *computer vision*. Berangkat dari hal tersebut, penelitian pemodelan 3D mulai diterapkan pada wajah manusia, sehingga memunculkan teknologi 3D *face reconstruction*. Pada sebuah penelitian menyatakan bahwa perkiraan rekonstruksi wajah 3D pada struktur geometris wajah tidak akan stabil untuk citra yang berbeda dari objek yang sama [2]. Maka dari itu, 3D *Morphable Face Models* (3DMM) dibuat dengan tujuan awal untuk mendukung *face recognition* dan 3D *mask printing* [2]. Akan tetapi, seiring dengan berkembangnya teknologi, 3DMM menjadi salah satu fundamental dalam penelitian di bidang *face alignment* dan 3D *face reconstruction*. Maka dari itu, perkembangan sistem 3D *face reconstruction* tidak lepas dari *face alignment*.

Beberapa penelitian *face alignment* dengan dataset AFLW2000-3D telah menunjukkan beberapa hasil yang mengejutkan. Pada penelitian Chandrasekhar diketahui bahwa dengan diterapkannya metode 3D *Spatial*

Transformer Network (3DSTN) menghasilkan rata - rata nilai *Normalized Mean Error* (NME) sebesar 4,49% [3]. Pada penelitian Xiangyu diketahui bahwa dengan metode *3D Dense Face Alignment* (3DDFA) menghasilkan rata – rata nilai NME sebesar 3,79% pada *face alignment*, dan sebesar 5,3695% pada proses *face reconstruction* [4]. Beberapa metode tersebut bekerja dengan baik pada wajah dengan pose frontal, namun tidak bisa mempertahankan nilai error yang rendah pada pose yang semakin kompleks [5]. Hal ini terjadi karena kinerja CNN pada pemodelan 3D terbatas karena adanya ruang 3D yang ditentukan oleh basis model wajah atau template [5]. Untuk mengatasi keterbatasan dari CNN, penelitian Yao Feng mengusulkan metode *Position Map Regression Network* (PRNet) [5]. Untuk pertama kalinya, metode ini secara bersamaan memecahkan masalah pada sistem *face alignment* dan 3D *face reconstruction* tanpa adanya batasan pada ruang solusi berdimensi rendah [5]. Pada metode PRNet dirancang representasi 2D yang disebut *UV position map* yang merekam bentuk 3D dari sebuah wajah pada ruang koordinat UV, dan dilakukan train sederhana pada CNN sehingga bisa mendapatkan informasi geometri hanya dari satu citra [5]. Integrasi *loss function* juga dilakukan untuk meningkatkan kinerja *network*. Penelitian ini memperoleh hasil rata – rata nilai NME sebesar 3,62% pada proses *face alignment* dan 3,9625% pada *face reconstruction* [5]. Berdasarkan data tersebut, proses *face reconstruction* dengan metode PRNet pada dataset AFLW2000-3D masih menjadi yang terbaik. Akan tetapi, apabila melihat dari kinerja *face alignment*, metode ini masih tidak sebanding dengan beberapa metode baru seperti 2DASL dan 3DDFA V2.

Berangkat dari permasalahan tersebut, pada Tugas Akhir ini diusulkan untuk melakukan pengujian *face alignment* pada metode PRNet dengan beberapa skema untuk mengetahui sampai sejauh mana metode ini dapat bekerja dengan baik. Skema pengujian dilakukan dengan melakukan perubahan mulai dari parameter jaringan dan juga hyperparameter, serta modifikasi arsitektur PRNet. Selain itu, dilakukan pengujian dengan melakukan pengelompokan citra uji untuk melihat seberapa jauh pengaruh kondisi citra pada performansi *face alignment*. Setelah melalui beberapa pengujian tersebut, dilakukan analisis untuk mendapatkan konfigurasi dengan performansi terbaik. Dengan ini diharapkan tugas akhir ini dapat digunakan sebagai referensi di penelitian selanjutnya.

2. Konsep Dasar

2.1. 3D Morphable Face Model

Pengenalan wajah otomatis dari satu gambar masih sulit untuk tampilan *non-frontal* dan kondisi pencahayaan yang kompleks [1]. Untuk mencapai pose dan variasi cahaya yang diinginkan, informasi 3D dari objek sangat dibutuhkan. Untuk alasan ini, *3D Morphable Face Models* (3DMM) diperkenalkan. 3DMM adalah model generatif bentuk dan tampilan wajah yang berdasar pada dua gagasan utama [6].

1. Semua wajah berada dalam korespondensi titik – ke – titik , yang biasanya ditetapkan pada sekumpulan contoh wajah dalam prosedur pendaftaran dan akan dipertahankan selama langkah pemrosesan selanjutnya [6]. Oleh karena itu, kombinasi linier wajah dapat didefinisikan dengan cara yang pasti, yang dimana akan menghasilkan wajah yang secara morfologis realistis (*morphs*) [6].
2. Memisahkan bentuk dan warna wajah seta menguraikannya dari faktor eksternal seperti pencahayaan dan parameter kamera [6].

2.2. Face Alignment

Teknologi penyalarsan wajah atau *face alignment* adalah teknologi *computer vision* untuk mengidentifikasi struktur geometris wajah manusia dalam sebuah citra digital [7]. Dengan kata lain, dapat diartikan sebagai teknologi untuk mengalokasikan serangkaian titik penting pada wajah manusia [8]. Pada proses penyalarsan wajah pada skala besar, dapat dibagi menjadi dua kategori yaitu *multi-view based* dan *3D model based*

2.3. Basel Face Model

Penggunaan 3DMM yang meluas seringkali terhambat oleh proses konstruksi yang sulit, yang membutuhkan pemindai 3D yang tepat dan cepat [6]. Hal ini pula yang semakin memperkuat bahwa satu – satunya kelemahan dari 3DMM adalah persyaratan dari model 3D [10]. *Basel Face Model* (BFM) model yang tidak hanya memungkinkan pengembangan algoritma analisis gambar berbasis 3DMM, tetapi juga akan memungkinkan praktik baru yang sebelumnya tidak mungkin untuk dilakukan, seperti kesulitan untuk menggeneralisasi dari satu *database* ke *database* lainnya, karena kondisi pencitraan yang terlalu berbeda [1]. Pada metode PRN sendiri, proses pembuatan UV position map diperoleh melalui *generate* data dari BFM *raw model* [1], informasi BFM tambahan dari penelitian sebelumnya seperti *Face Profiling* dan 3DDFA [4], informasi HPEN [11], serta koordinat UV BFM [3].

2.4. Convolutional Neural Networks

Convolutional Neural Networks (CNN) merupakan salah satu *neural networks* yang paling populer digunakan, terutama pada data berdimensi tinggi, seperti citra dan video [12]. CNN berfungsi untuk klasifikasi citra, segmentasi semantik, deteksi objek dan ekstraksi ciri [12]. CNN bekerja dengan operasi konvolusi. Secara umum, CNN terbagi menjadi tiga *layer*, di antaranya *convolutional layer*, *pooling layer* dan *fully-connected layer* serta terdiri dari neuron yang memiliki *weight*, *bias* dan *activation function* [13]. Pada metode PRNet, digunakan

arsitektur berbasis *encoder – decoder* yang terdiri dari *convolutional layer*, *transposed convolutional layer*, dan *ReLU* sebagai aktivasi serta *residual blocks* sebagai *backbone*.

2.5. Residual Blocks

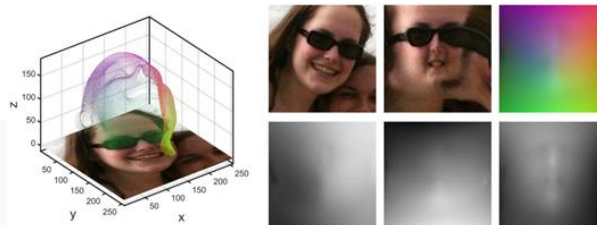
Arsitektur CNN dengan kedalaman jaringan yang tinggi menciptakan kinerja yang baik, namun juga menimbulkan beberapa masalah, salah satunya adalah *vanishing gradient*, yaitu kondisi dimana gradien hasil *training* tidak dapat mencapai *layer* pertama karena mengalami konvolusi berkali – kali sehingga pada *layer* pertama tidak mendapatkan gradien apapun. Hal ini akan menyebabkan CNN tidak dapat mendapatkan informasi dari error yang telah dikalkulasi. Ini terjadi karena setiap *layer* hanya diumpangkan pada *layer* selanjutnya. Maka dari itu, pada jaringan dengan *residual blocks*, setiap *layer* akan diumpangkan ke *layer* berikutnya serta ke 2 sampai 3 *layer* di depan dengan mengandalkan jalan pintas (*short-cut*) [14].

2.6. UV Position Map

UV position map atau peta posisi merupakan gambar 2D yang merekam posisi 3D dari semua titik dalam ruang UV. Dengan kata lain, peta posisi dapat dengan mudah dipahami yaitu dengan mengganti nilai RGB pada peta tekstur dengan koordinat *x*, *y*, *z* [5]. Dengan demikian, *position map* dapat dijabarkan dengan persamaan 1,

$$Pos(u_i, v_i) = (x_i, y_i, z_i) \tag{1}$$

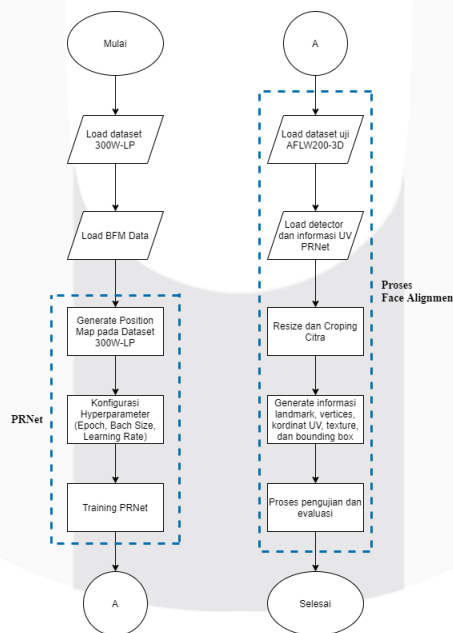
Dengan (u_i, v_i) merepresentasikan koordinat UV dari sebuah permukaan dan (x_i, y_i, z_i) merepresentasikan posisi 3D yang sesuai dari struktur wajah, dengan x_i dan y_i sebagai posisi 2D yang sesuai dari wajah pada citra RGB *input*, serta z_i sebagai kedalaman bidang, seperti yang ditunjukkan pada Gambar 1.



Gambar 1. Ilustrasi UV Position Map [5].

3. Desain Model Sistem

Pada Tugas Akhir ini dilakukan analisis performa pada proses *face alignment*. Skema dari Tugas Akhir ini menggunakan metode PRNet. Secara umum Tugas Akhir ini memiliki alur kerja seperti pada Gambar 2.



Gambar 2. Diagram Alir Sistem

3.1. Dataset 300W-LP

Pada Tugas Akhir ini menggunakan dataset 300W-LP sebagai dataset latih. Dataset 300W-LP merupakan dataset *open-source* yang terdiri dari sekumpulan citra wajah dengan pose bervariasi yang merupakan hasil sintesis dari dataset 300W. Dataset 300W-LP terdiri dari 122.450 citra wajah. Dataset ini digunakan karena sudah terdapat informasi geometri yang sudah sesuai dengan wajah 2D serta 3D, sehingga tidak perlu dilakukan proses *generate* informasi yang salah satu caranya biasa dilakukan melalui metode *Iterative Closest Point* (ICP) untuk mendapatkan informasi tersebut.

3.2. Dataset AFLW2000-3D

Pada proses pengujian *face alignment* digunakan dataset AFLW2000-3D. Dataset AFLW2000-3D terdiri dari 2.000 citra wajah dengan informasi *landmark* 3D, sehingga dataset ini tidak hanya dapat digunakan untuk proses evaluasi pada 2D, tapi juga pada 3D *face alignment*. Pada dataset ini terdiri citra wajah antara sudut -90° sampai dengan 90° . Pada proses pengujian dan analisis dilakukan skema pengelompokan dataset. Dataset dikelompokkan berdasarkan pose wajah dan *yaw angle*. Detail pengelompokan dataset berdasarkan pose wajah dapat dilihat pada Tabel 1.

Tabel 1. Pengelompokan citra uji

	Pengelompokan	Jumlah Citra	Total
Pose Wajah	Depan	1168	2000
	Kanan	417	
	Kiri	410	
	Kompleks	9	
Yaw Angle	$[0^\circ, 30^\circ]$	1239	2000
	$[30^\circ, 60^\circ]$	304	
	$[60^\circ, 90^\circ]$	457	

3.3. Load BFM Data

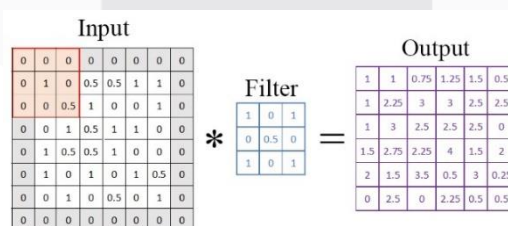
Sebelum melakukan *generate UV Position Map*, dibutuhkan beberapa informasi data tambahan yang didapat dari BFM dan beberapa penelitian sebelumnya. Informasi data tambahan tersebut diantaranya adalah BFM *raw model*, *face profiling*, Algoritma 3DDFA, HPEN, dan koordinat UV BFM. BFM *raw model* merupakan algoritma dasar BFM yang digunakan sebagai *template* informasi untuk menampung informasi berdasarkan parameter BFM [1]. *Face profiling* berisi informasi untuk melakukan sintesis pose wajah dengan model wajah 3D [4]. Algoritma 3DDFA berfungsi untuk menyesuaikan parameter 3DMM pada berbagai pose citra wajah [4]. HPEN berfungsi sebagai normalisasi pose wajah dan ekspresi [11]. Koordinat UV BFM berisi informasi 3D untuk mengubah citra kedalam koordinat UV berdasarkan parameter BFM [3].

3.4. Konfigurasi Sistem

Pada Tugas Akhir ini menggunakan skema konfigurasi *hyperparameter* diantaranya berupa *epoch*, *batch size*, dan *learning rate*. Tugas Akhir ini menggunakan konfigurasi *epoch* sebesar 50, dengan konfigurasi *batch size* 16 dan *learning rate* pada 0.0001. Hal ini dilakukan karena mengacu pada penelitian Yao Feng [5], disebutkan bahwa konfigurasi terbaik ada pada *batch size* 16 dan *learning rate* 0,0001. Selain itu, digunakan dua macam skema konfigurasi, yaitu skema *padding* dan skema modifikasi.

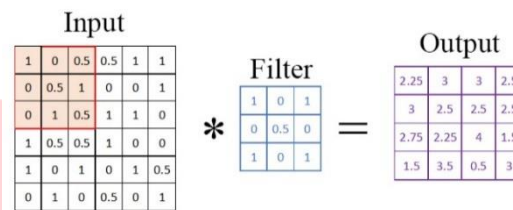
3.5. Skema Padding

Pada skema *padding*, digunakan dua macam skema konfigurasi, yaitu skema *same-padding* dan *valid-padding*. Skema *same-padding* menghasilkan output *feature map* sebesar dimensi citra input. Dengan kata lain, jika citra masukan memiliki dimensi 6×6 , maka keluaran yang dihasilkan apabila menggunakan skema *same-padding* akan sama dengan dimensi masukan, yaitu 6×6 . Proses skema *same-padding* pada sebuah layer diilustrasikan pada Gambar 3.



Gambar 3. Ilustrasi skema *same-padding*.

Pada skema *valid-padding* tidak ada penambahan *padding* seperti pada *same-padding*, sehingga dimensi keluaran dari proses konvolusi akan mengalami reduksi. Artinya, *feature map* yang dihasilkan pada *valid-padding* lebih sedikit dibandingkan *same-padding* karena adanya reduksi dimensi. Proses skema *valid-padding* pada sebuah *layer* diilustrasikan pada Gambar 4.



Gambar 4. Ilustrasi skema *valid-padding*.

3.6. Skema Modifikasi

Selain menerapkan dua skema padding, pada Tugas Akhir ini dilakukan juga modifikasi pada konfigurasi arsitektur asli PRNet. Pada skema modifikasi ini, dibuat dua opsi modifikasi konfigurasi untuk melihat bagaimana performansi yang dihasilkan dan dibandingkan dengan skema asli PRNet. Pada skema modifikasi pertama, ditambahkan satu *layer* konvolusi pada *layer* pertama dengan parameter yang sama dengan *layer* kedua. Tujuan dari konfigurasi ini adalah untuk mencoba melihat bagaimana pengaruh penambahan *layer* pada konfigurasi asli PRNet. Akan tetapi dengan diterapkannya parameter yang sama dengan *layer* kedua, maka terjadi adanya pengurangan dimensi pada peta fitur sehingga terjadi hilangnya beberapa informasi. Rincian konfigurasi *layer* yang digunakan dapat dilihat pada Tabel 2.

Tabel 2. Konfigurasi layer Modifikasi 1.

Layer name	Layer
Conv_layer	$4 \times 4, 3$ $4 \times 4, 16$
Resblock	$\begin{bmatrix} 4 \times 4, 32 \\ 4 \times 4, 64 \\ 4 \times 4, 128 \\ 4 \times 4, 256 \\ 4 \times 4, 512 \end{bmatrix} \times 2$
Transposedconv_layer	$4 \times 4, 512$
	$\begin{bmatrix} 4 \times 4, 256 \\ 4 \times 4, 128 \\ 4 \times 4, 64 \end{bmatrix} \times 3$
	$\begin{bmatrix} 4 \times 4, 32 \\ 4 \times 4, 16 \end{bmatrix} \times 2$
	$[4 \times 4, 3] \times 3$

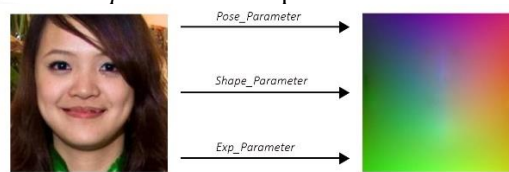
Pada skema ini, ditambahkan juga satu *layer* konvolusi pada *layer* pertama, namun dengan menerapkan skema *same-padding* pada *layer* yang ditambahkan. Dengan dilakukannya penambahan *padding*, informasi citra keluaran dari *layer* pertama tidak akan mengalami pengurangan dimensi. Oleh karena itu, dengan menerapkan skema ini, informasi yang didapatkan akan lebih jauh diolah tanpa terjadinya pengurangan informasi yang akan diteruskan ke *layer* selanjutnya. Rincian konfigurasi *layer* yang digunakan dapat dilihat pada Tabel 3.

Tabel 3. Konfigurasi layer Modifikasi 2

Layer name	Layer
Conv_layer	$3 \times 3, 3$ (<i>same-padding</i>) $4 \times 4, 16$
Resblock	$\begin{bmatrix} 4 \times 4, 32 \\ 4 \times 4, 64 \\ 4 \times 4, 128 \\ 4 \times 4, 256 \\ 4 \times 4, 512 \end{bmatrix} \times 2$
Transposedconv_layer	$4 \times 4, 512$
	$\begin{bmatrix} 4 \times 4, 256 \\ 4 \times 4, 128 \\ 4 \times 4, 64 \end{bmatrix} \times 3$
	$\begin{bmatrix} 4 \times 4, 32 \\ 4 \times 4, 16 \end{bmatrix} \times 2$
	$[4 \times 4, 3] \times 3$

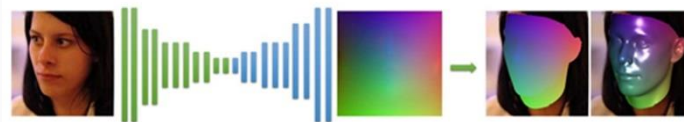
3.7. Load UV Position Map

Sebelum melakukan proses *training*, dataset latih terlebih dahulu diubah menjadi koordinat UV dengan nama UV *position map*. UV *Position Map* pada dasarnya tercipta berdasarkan informasi parameter suatu citra, yaitu parameter pose, parameter bentuk (*shape*), dan parameter ekspresi (*expression*) yang diubah dan dipetakan menjadi informasi 3D berdasarkan parameter koordinat UV BFM. Parameter tersebut melalui proses transformasi *mesh* dan *cropping* dengan *keypoint* sebelum nantinya dilakukan perubahan pada posisi wajah dengan mengubah skala dan melakukan translasi pada sumbu *z* hingga akhirnya diubah menjadi koordinat UV berbentuk RGB. Proses pengubahan citra menjadi UV *position map* diilustrasikan pada Gambar 5.

Gambar 5. Proses generate parameter UV *position map*.

3.7. Train PRNet

Pada proses *training*, dilakukan *cropping* pada citra input sesuai dengan *ground truth bounding box* dan memperkecil dimensinya menjadi 256×256 . Kemudian, dengan memanfaatkan parameter yang berdasar pada 3DMM dapat menghasilkan posisi 3D yang sesuai, dan mengubahnya menjadi ruang UV untuk memperoleh informasi *ground truth position map* dengan ukuran yang sama, yaitu 256×256 . Perlu diketahui, meskipun data latih dihasilkan melalui informasi 3DMM, *position map* hasil keluaran dari jaringan PRNet tidak terbatas hanya pada *template* ataupun ruang linier 3DMM apapun. Pada arsitektur PRNet, digunakan arsitektur berbasis *encoder* dan *decoder* yang terdiri dari *convolutional layer*, *transposed convolutional layer*, *residual blocks* dan *ReLU layer* sebagai aktivasi.



Gambar 6. Arsitektur jaringan PRNet [5].

Seperti yang ditunjukkan pada Gambar 6, persegi panjang berwarna hijau merepresentasikan *residual blocks*, dan persegi panjang berwarna biru merepresentasikan *transposed convolutional layer*. Pada sisi *encoder* terdiri dari *convolutional layer* diikuti dengan sepuluh *residual blocks*. Sedangkan pada sisi *decoder* terdiri dari tujuh belas *transposed convolutional layer*.

Untuk mengukur seberapa besar prediksi *error* dan meningkatkan performansi jaringan, dilakukan integrasi *loss function*. *Loss function* pada PRNet diperoleh melalui integrasi dari *weight mask*. *Weight mask* adalah citra abu-abu yang merekam kedalaman setiap titik pada *position map*. Ilustrasi *weight mask* ditunjukkan pada Gambar 7.



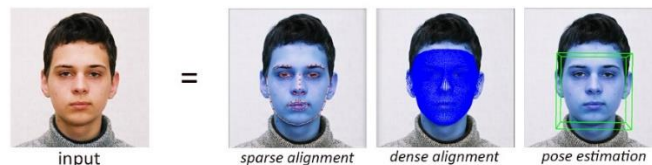
Gambar 7. Weight mask [5].

Sebagaimana diketahui sebelumnya, $Pos(u, v)$ merepresentasikan *position map* hasil prediksi dengan (u, v) yang merupakan koordinat masing – masing *pixel*. *Ground truth position map* direpresentasikan dengan $\widehat{Pos}(u, v)$ dan $W(u, v)$ adalah *weight mask*. Oleh karena itu, *loss function* yang digunakan didefinisikan dengan persamaan 2 [5].

$$Loss = \sum \|Pos(u, v) - \widehat{Pos}(u, v)\| \cdot W(u, v) \quad (2)$$

3.8. Proses Face Alignment

Informasi yang dihasilkan dari *training model* adalah *landmarks*, simpul (*vertices*), *texture*, dan warna. Pada informasi *landmarks* akan diperoleh beberapa parameter, salah satunya adalah *keypoint*, sedangkan pada *vertices* terdapat informasi *camera matrix*. Proses *face alignment* terlebih dahulu melewati proses pendeteksian wajah menggunakan *library dlib*. *Dlib* akan melakukan deteksi wajah berdasarkan informasi yang didapat dari *training model*.



Gambar 8. Ilustrasi face alignment.

Pada Gambar 8 menunjukkan bahwa *face alignment* dibagi menjadi tiga jenis berdasarkan informasi yang diproses, yaitu *sparse alignment*, *dense alignment*, dan *pose estimation*. *Sparse alignment* merupakan alignment yang diperoleh dari informasi *keypoint* dan *landmarks* wajah, pada kasus ini *sparse alignment* merekam 68 *landmark* wajah. *Dense alignment* diperoleh dari informasi *vertices*, pada kasus ini *dense alignment* merekam informasi sebanyak 45K *point* wajah. *Pose estimation* diperoleh berdasarkan informasi *keypoint* dan *camera matrix*.

3.9. Parameter Performansi

Secara metode tradisional, metrik evaluasi untuk *face alignment* adalah dengan menghitung jarak *Euclidean* dari titik ke titik yang dinormalisasi oleh jarak interokuler. Akan tetapi, metrik ini bersifat bias pada beberapa citra wajah yang jarak interokulernya sangat kecil. Oleh karena itu, pada Tugas Akhir ini dilakukan normalisasi antara *landmark* wajah dengan *bounding box*. Secara khusus, digunakan parameter performansi *Normalized Mean Error* (NME) yang didefinisikan dengan persamaan 3.2 [15],

$$NME = \frac{1}{N} \sum_{k=1}^N \frac{\|x_k - y_k\|^2}{d} \quad (3)$$

dimana x merepresentasikan *landmark ground truth* dari wajah yang diketahui, y merupakan nilai hasil prediksi, dan N merupakan banyaknya citra yang mengalami pengujian. Sebagai faktor normalisasi, d adalah akar dari luas *bounding box* yang diperoleh dari w_{bbox} sebagai lebar *bounding box* dikali dengan h_{bbox} yang merupakan tinggi dari *bounding box*, sebagaimana didefinisikan pada persamaan 3.

$$d = \sqrt{w_{bbox} \times h_{bbox}} \quad (4)$$

4. Hasil dan Analisis

Bagian ini berisi tentang hasil dan analisis dari pengujian. Dalam bab ini dilakukan pengujian pada beberapa skema konfigurasi dan pengelompokan citra. Dalam bab ini juga didapatkan konfigurasi terbaik pada sistem *face alignment* dengan metode PRNet

4.1. Analisis Pengujian Berdasarkan Pose Wajah

Hasil pengujian pada citra yang dikelompokkan berdasarkan pose wajah seperti yang ditunjukkan pada Tabel 4 menunjukkan bahwa konfigurasi modifikasi menjadi konfigurasi dengan nilai rata - rata terbaik. Perbedaan kinerja setiap konfigurasi pada setiap pose wajah disebabkan karena beberapa faktor, yaitu jenis citra yang lebih terstruktur karena adanya pengelompokan dan kecocokan citra tersebut dengan konfigurasi arsitektur yang digunakan. Pada konfigurasi *same-padding*, pose wajah tampak depan, kanan, dan kiri memiliki performansi yang lebih buruk apabila dengan konfigurasi *valid-padding*. Disisi lain, pada pose wajah kompleks, konfigurasi *same-padding* jauh lebih unggul dibandingkan konfigurasi *valid-padding*, hal ini disebabkan karena penerapan padding lebih cocok digunakan untuk jenis citra pada kelompok pose wajah kompleks.

Tabel 4. Pengujian berdasarkan pose.

Titik Penyelarasan	Konfigurasi	NME 2D/3D	Pengelompokan Pose				Mean	
			Depan	Kanan	Kiri	Kompleks		
68 Landmark	Original PRNet	2D	3.12	4.31	4.43	10.89	5.69	
		3D	4.39	5.86	6.09	13.95	7.58	
	Modifikasi	2D	2.99	4.26	4.63	9.91	5.45	
		3D	4.34	5.86	6.24	13.48	7.48	
	Same-Padding	2D	3.19	4.39	4.59	14.02	6.55	
		3D	4.48	6.08	6.26	16.96	8.44	
	Valid-Padding	2D	3.15	4.38	4.61	15.52	6.91	
		3D	4.40	5.98	6.24	19.14	8.94	
	45K Point	Original PRNet	2D	2.76	4.27	4.26	10.34	5.41
			3D	3.86	5.61	5.77	12.97	7.05
Modifikasi		2D	2.73	4.18	4.43	8.77	5.03	
		3D	3.87	5.60	5.90	12.66	7.01	
Same-Padding		2D	2.81	4.30	4.42	11.30	5.71	
		3D	3.93	5.76	5.92	14.52	7.53	
Valid-Padding		2D	2.78	4.29	4.40	13.64	6.28	
		3D	3.89	5.74	5.93	17.52	8.27	

4.2. Analisis Pengujian Berdasarkan Pose Wajah

Hasil pengujian pada citra yang dikelompokkan berdasarkan *yaw angle* seperti yang ditunjukkan pada Tabel 5 menunjukkan bahwa konfigurasi original dan modifikasi memiliki performansi yang sama khususnya untuk NME 2D pada 64 landmark wajah, dan mampu menyaingi konfigurasi asli pada *yaw angle* lainnya. Perbedaan kinerja setiap konfigurasi pada setiap pose wajah disebabkan karena beberapa faktor, yaitu jenis citra yang lebih terstruktur karena adanya pengelompokan dan kecocokan citra tersebut dengan konfigurasi arsitektur yang digunakan. Pada pengujian *yaw angle* [30°, 60°], konfigurasi *valid-padding* memiliki performansi terbaik dibandingkan konfigurasi lainnya. Hal ini dikarenakan konfigurasi *kernel* yang lebih kecil tanpa penambahan padding lebih cocok digunakan pada kelompok sudut tersebut.

Tabel 5. Pengujian berdasarkan *yaw angle*.

Titik Penyelarasan	Konfigurasi	NME 2D/3D	Yaw Angle				
			[0,30]	[30,60]	[60,90]	Mean	
68 Landmark	Original PRNet	2D	3.14	3.71	5.05	3.97	
		3D	4.44	5.14	6.76	5.45	
	Modifikasi	2D	3.04	3.69	5.19	3.97	
		3D	4.42	5.15	6.95	5.51	
	Same-Padding	2D	3.20	3.73	5.28	4.07	
		3D	4.53	5.21	7.12	5.62	
	Valid-Padding	2D	3.19	3.67	5.30	4.05	
		3D	4.49	5.09	7.07	5.55	
	45K Point	Original PRNet	2D	2.79	3.45	4.98	3.74
			3D	3.93	4.76	6.46	5.05
Modifikasi		2D	2.79	3.45	5.07	3.77	
		3D	3.97	4.81	6.65	5.14	
Same-Padding		2D	2.84	3.48	5.17	3.83	
		3D	3.99	4.81	6.79	5.20	
Valid-Padding		2D	2.85	3.41	5.20	3.82	
		3D	4.00	4.71	6.82	5.18	

4. Kesimpulan

Sistem *face alignment* dengan menggunakan metode PRNet telah berhasil dilakukan pengujian dan analisis. Modifikasi pada konfigurasi juga berhasil diimplementasikan. Secara keseluruhan, performansi terbaik NME 2D yang didapatkan pada konfigurasi Modifikasi dengan nilai terbaik pada 68 landmark dan 45 ribu point masing - masing sebesar 3.97% dan 3.77%. Disisi lain, performansi terbaik pada NME 2D yang didapatkan pada konfigurasi asli PRNet dengan nilai terbaik pada 68 landmark dan 45 ribu point masing - masing sebesar 3.87% dan 3.77%. Untuk NME 3D pada konfigurasi modifikasi didapatkan nilai sebesar 5.51% untuk 68 landmark 5.14% untuk 45k point, sedangkan pada konfigurasi original diperoleh nilai sebesar 5.45% untuk 68 landmark dan 5.14% untuk 45k

point. Penerapan skema *same-padding* kurang berpengaruh pada peningkatan performansi, karena jenis citra yang diuji sangat berpengaruh pada performansi konfigurasi. Penggunaan ukuran *kernel* yang lebih kecil membuat pengenalan gambar lebih detail sehingga konfigurasi terbaik relatif didapat dengan nilai *epoch* yang lebih kecil.

5. Referensi

- [1] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, dan T. Vetter, "A 3D face model for pose and illumination invariant face recognition," *6th IEEE Int. Conf. Adv. Video Signal Based Surveillance, AVSS 2009*, pp. 296–301, 2009.
- [2] L. Fangmin, C. Ke, dan L. Xinhua, "3D Face Reconstruction Based on Convolutional Neural Network," *Proc. - 10th Int. Conf. Intell. Comput. Technol. Autom. ICICTA 2017*, vol. 2017-October, no. 2, pp. 71–74, 2017.
- [3] C. Bhagavatula, C. Zhu, K. Luu, dan M. Savvides, "Faster than Real-Time Facial Alignment: A 3D Spatial Transformer Network Approach in Unconstrained Poses," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2017-October, pp. 4000–4009, 2017.
- [4] X. Zhu, X. Liu, Z. Lei, dan S. Z. Li, "Face Alignment in Full Pose Range: A 3D Total Solution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 1, pp. 78–92, 2019.
- [5] Y. Feng, F. Wu, X. Shao, Y. Wang, dan X. Zhou, "Joint 3d face reconstruction and dense alignment with position map regression network," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11218 LNCS, pp. 557–574, 2018.
- [6] B. Egger *et al.*, "3D Morphable Face Models—Past, Present, and Future," *ACM Trans. Graph.*, vol. 39, no. 5, 2020.
- [7] L. Gu dan T. Kanade, "Face Alignment BT - Encyclopedia of Biometrics," S. Z. Li and A. Jain, Eds. Boston, MA: Springer US, 2009, pp. 291–294.
- [8] A. Bulat dan G. Tzimiropoulos, "Convolutional aggregation of local evidence for large pose face alignment," *Br. Mach. Vis. Conf. 2016, BMVC 2016*, vol. 2016-Sept, pp. 86.1-86.12, 2016.
- [9] F. Liu, Q. Zhao, X. Liu, dan D. Zeng, "Joint Face Alignment and 3D Face Reconstruction with Application to Face Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 3, pp. 664–678, 2020.
- [10] S. Kevin Zhou dan R. Chellappa, "Illuminating Light Field: Image-based Face Recognition Across Illuminations and Poses Shaohua Kevin Zhou and Rama Chellappa Center for Automation Research and ECE Department University of Maryland, College Park, MD 20742," *Gesture*, pp. 10–15, 2004.
- [11] X. Zhu, Z. Lei, J. Yan, D. Yi, dan S. Z. Li, "High-fidelity Pose and Expression Normalization for face recognition in the wild," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07-12-June, pp. 787–796, 2015.
- [12] S. Khan, H. Rahmani, S. A. A. Shah, dan M. Bennamoun, "A Guide to Convolutional Neural Networks for Computer Vision," *Synth. Lect. Comput. Vis.*, vol. 8, no. 1, pp. 1–207, 2018.
- [13] M. Coskun, A. Ucar, O. Yildirim, dan Y. Demir, "Face Recognition Based on Convolutional Neural Network," *2017 Int. Conf. Mod. Electr. Energy Syst.*, pp. 376–379, 2017.
- [14] S. Sahoo, "Residual blocks — Building blocks of ResNet," 2018. <https://towardsdatascience.com/residual-blocks-building-blocks-of-resnet-fd90ca15d6ec> (accessed Oct. 19, 2020).
- [15] A. Bulat dan G. Tzimiropoulos, "How far are we from solving the 2D and 3D Face Alignment problem? (and a dataset of 230,000 3D facial