

# **Judul Proyek: Real-Time 3D Facial Expression Retargeting Using Dense Landmark-Based Regression**

Tim Mahasiswa:

Bryan Austin Hidayat / 2702357384

Bryan Santosa / 2702351052

Kinsley Reynard Tanjung / 2702349432

---

## **I. Introduction (Pendahuluan)**

### **1.1. Latar Belakang**

Perkembangan teknologi visi komputer telah memungkinkan interaksi manusia-komputer melalui pemrosesan wajah secara real-time, seperti filter augmented reality di media sosial, avatar 3D interaktif, dan sistem VTuber. Aplikasi-aplikasi tersebut membutuhkan kemampuan untuk mendeteksi wajah, mengekstraksi landmark, dan memprediksi ekspresi pengguna dengan cepat, stabil, dan responsif terhadap perubahan pose maupun pencahayaan.

Pendekatan populer dalam tugas ini adalah regresi 3D Morphable Model (3DMM) menggunakan jaringan syaraf konvolusional (CNN). Salah satu contohnya adalah 3DDFA\_v2 yang memanfaatkan FaceBoxes sebagai face detector dan CNN sebagai backbone untuk memproyeksikan citra 2D menjadi parameter wajah 3D. Walaupun metode ini menghasilkan estimasi ekspresi dan bentuk wajah secara akurat, pendekatan tersebut relatif berat secara komputasi dan sensitif terhadap kualitas input. Pada kondisi yang tidak ideal—misalnya pencahayaan rendah, pose ekstrem, atau occlusion seperti kacamata—pipeline berbasis CNN dapat menghasilkan landmark yang tidak stabil dan ekspresi yang tidak konsisten.

Sebagai alternatif, muncul metode yang lebih ringan seperti MediaPipe FaceMesh yang secara langsung memprediksi ratusan landmark wajah 3D secara real-time bahkan pada perangkat mobile. Framework ini menawarkan robustness lebih tinggi pada variasi pose, serta tidak bergantung pada running biaya inference CNN yang besar. Dengan memanfaatkan FaceMesh sebagai pengganti CNN + FaceBoxes, dan menggabungkannya dengan model neural network ringan seperti Multilayer

Perceptron (MLP), dimungkinkan untuk menurunkan kompleksitas sistem sambil tetap mempertahankan akurasi estimasi ekspresi dari wajah pengguna.

Berangkat dari kebutuhan tersebut, proyek ini mengeksplorasi pipeline alternatif berbasis MediaPipe FaceMesh + MLP untuk menggantikan CNN + FaceBoxes pada 3DDFA, serta mengevaluasi seberapa dekat hasil prediksi ekspresi model ringan dibanding ground truth ekspresi pada dataset wajah.

## 1.2. Tujuan

Penelitian ini bertujuan untuk mengevaluasi dan membandingkan efektivitas pipeline baseline 3DDFA\_v2 berbasis CNN dengan pipeline alternatif yang lebih ringan berbasis MediaPipe FaceMesh dan Multilayer Perceptron (MLP) dalam memprediksi parameter wajah 3D dan ekspresi. Melalui serangkaian eksperimen, penelitian ini berfokus pada pengujian kemampuan model ringan dalam menghasilkan parameter ekspresi yang mendekati ground truth, menilai performa sistem pada kondisi real-time dan skenario dunia nyata, serta mengevaluasi efisiensi dan akurasi pendekatan landmark-based regression dibandingkan model CNN yang lebih kompleks. Selain itu, penelitian ini juga bertujuan untuk mengukur kualitas prediksi secara kuantitatif dan kualitatif, serta membuktikan kelayakan pipeline yang diusulkan sebagai solusi yang lebih efisien dan siap digunakan dalam sistem produksi berbasis estimasi ekspresi wajah 3D.

## 1.3. Ruang Lingkup

Ruang lingkup penelitian ini mencakup keseluruhan proses pengembangan sistem real-time facial expression retargeting berbasis deep learning, dimulai dari eksplorasi dataset hingga deployment aplikasi. Tahapan awal meliputi analisis dan eksplorasi dataset AFLW2000-3D untuk memahami struktur data landmark dan ekspresi, diikuti proses preprocessing, normalisasi landmark, serta penyiapan ground truth ekspresi sebagai dasar pelatihan. Penelitian berfokus pada pengembangan dan pembandingan dua pipeline inferensi wajah, yaitu baseline 3DDFA\_v2 yang menggunakan FaceBoxes dan CNN backbone, dengan pipeline alternatif yang diusulkan menggunakan MediaPipe FaceMesh sebagai landmark extractor dan Multilayer Perceptron (MLP) sebagai model regresi ekspresi. Pelatihan hanya dilakukan pada model MLP dengan menggunakan landmark hasil preprocessing sebagai input dan ekspresi dataset sebagai target. Evaluasi performa dilakukan secara kuantitatif menggunakan metrik MSE, MAE, dan NME, serta kualitatif melalui visualisasi mesh dan pengujian real-time via webcam. Ruang lingkup juga mencakup implementasi sistem inferensi berbasis web, yang terdiri dari frontend (Next.js + Three.js) untuk visualisasi 3D, backend inference service dengan FastAPI dan PyTorch untuk menjalankan model, serta deployment menggunakan Docker dan Azure Container Apps. Penelitian ini tidak mencakup pengembangan dataset baru, pelatihan ulang CNN, atau estimasi tekstur wajah; fokus dibatasi pada regresi

ekspresi, stabilitas prediksi landmark, dan demonstrasi sistem end-to-end secara real-time.

## II. Dataset

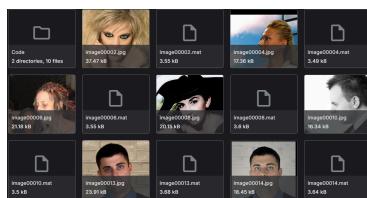
### 2.1. Data Explanation

Dataset yang kami gunakan itu berasal dari dataset AFLW2000-3D yang berisi 2 tipe data. Dataset tersebut berisi:

1. 2000 face images (jpg)
2. 2000 annotation (.mat)

Dataset tersebut berisikan gambar face yang akan digunakan untuk model yang kami gunakan, dataset tersebut berisi 2000 gambar images yang diambil dari beberapa angle yang berbeda agar model dapat mengetahui citra 2D dari wajah, ekspresi dan occlusion. Untuk yang file annotation, dia itu lebih ke memberi arah ke landmark facenya dalam bentuk 2D atau 3D, 3DMM parameter.

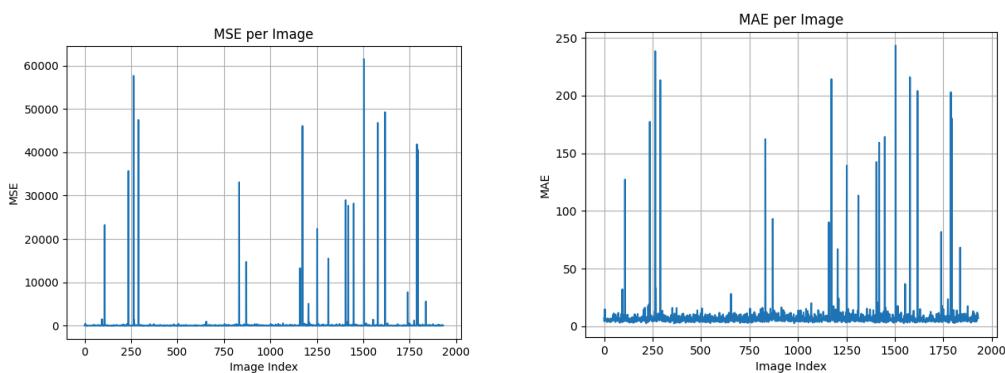
Gambaran umum untuk dataset yang digunakan:



Link dataset yang kami gunakan:

<https://www.kaggle.com/datasets/mohamedadlyi/aflw2000-3d>

### 2.2. Eksplorasi (EDA)



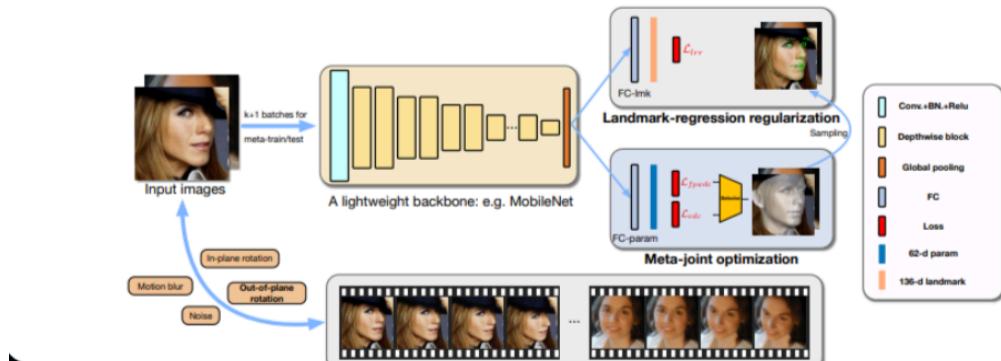
Dalam pengujian model Faceboxes + CNN pada dataset AFLW2000, kami menemukan bahwa performa model cenderung stabil pada mayoritas data, namun menunjukkan fluktuasi yang sangat ekstrem pada sampel tertentu. Hal ini terlihat dari grafik MSE yang melonjak hingga 60.000 dan MAE yang melebihi 200 pada beberapa indeks gambar. Tingginya lonjakan error ini mengindikasikan bahwa meskipun arsitektur backbone saat ini sudah akurat untuk kondisi wajah normal,

model masih mengalami kesulitan dalam menangani edge cases seperti oklusi atau sudut wajah yang sangat ekstrem (large head poses) yang menjadi karakteristik utama dataset AFLW2000. Temuan ini menjadi dasar bagi kami untuk mengevaluasi kembali arsitektur backbone guna meningkatkan ketahanan model terhadap variasi pose yang kompleks.

### III. Modeling (Pemodelan)

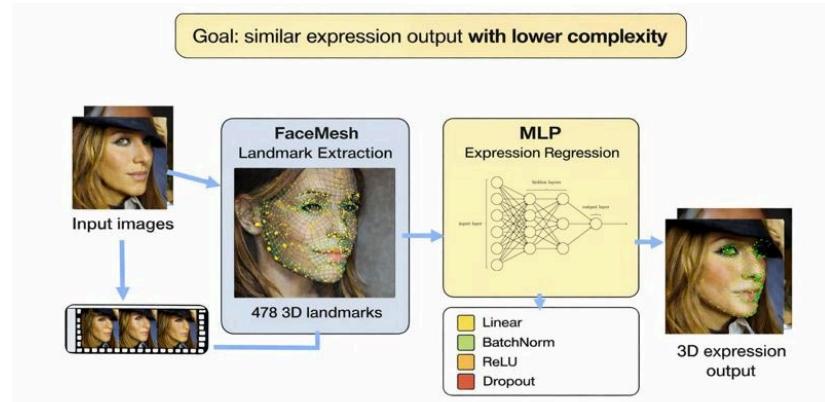
#### 3.1. Arsitektur Model:

Pada penelitian ini menggunakan 3DDFAv2 sebagai baseline model untuk permasalahan yang dihadapi contohnya pada regresi wajah 3D pada project kami. Model 3DDFAv2 ini kami gunakan karena model tersebut mampu meregresikan parameter 3D dengan menggunakan 3DMM (3D Morphable Model) yang dapat mencakup pose, shape, dan expression wajah, serta menghasilkan face landmark secara padat. Dari references yang kami ambil dari github cleardusk, arsitektur dari 3DDFAv2 menggunakan Faceboxes sebagai face detector dan menggunakan CNN sebagai backbone untuk ekstraksi visual fitur dan citra wajah. Namun, pendekatan berbasis CNN pada 3DDFAv2 itu memerlukan proses feature extraction yang based on pixel yang complex dan sensitif terhadap kualitas citra hasil deteksi wajah. Contohnya seperti jika ada occlusion, illumination change dan pose variance maka model tersebut seringkali miss predict. Gambar dibawah ini merupakan gambaran dari arsitektur 3DDFAv2 baseline yang menggunakan Faceboxes + CNN backbone.



Maka dari itu kami mencari alternatif pipeline yang bertujuan untuk menyederhanakan proses downstream regression, bukan untuk menggantikan keseluruhan dari pipeline tersebut. Pendekatan yang kami lakukan itu adalah dengan menggunakan Mediapipe Facemesh sebagai main module kami untuk mengeskstraksi 3D face landmark. Mediapipe facemesh ini akan melakukan face detection and 3D landmark regression, sehingga keluaran yang dihasilkan berupa representasi face geometry dalam bentuk 3D landmark yang terstruktur. Kemudian landmark yang dihasilkan dari mediapipe facemesh akan digunakan sebagai input untuk MLP (Multi Layer Perceptron) yang memiliki peran sebagai expression regression head. Dengan menggunakan pendekatan ini, proses visual feature extraction berbasis CNN pada

tahap downstream dapat dieliminasi, dan model regresi hanya perlu mempelajari hubungan geometris antar titik face landmark. Hal ini membuat pipeline akan menjadi lebih ringan dan terfokus pada expression regression, tanpa ketergantungan langsung pada raw pixel feature.



### 3.2. Skenario Pelatihan:

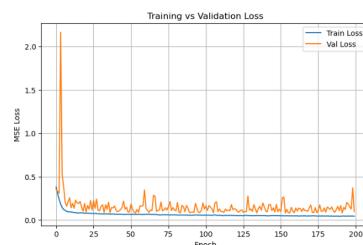
Pada eksperimen baseline, model 3DDFAv2 (FaceBoxes + CNN) ini digunakan dalam kondisi sudah pretrained, tanpa dilakukan pelatihan ulang pada backbone CNN maupun face detector. Model hanya dievaluasi untuk melihat performanya pada dataset yang digunakan dalam eksperimen kami.

Untuk bagian yang menggunakan Faceboxes + CNN, kami menggunakan model pretrained yang kami ambil dari repository resmi sehingga pada penelitian yang kami lakukan, kami tidak melakukan training dari awal pada bagian faceboxes + cnn melainkan hanya evaluasi performance dari model pada dataset yang kami gunakan.

Untuk model yang kami gunakan (Mediapipe facemesh + MLP), kami melakukan proses fine-tuning dengan konfigurasi sebagai berikut:

1. Optimizer: Adam
2. Epochs: 200
3. Learning rate: 1e-4
4. Weight decay :1e-5
5. Batch size: 128
6. Loss function: mse

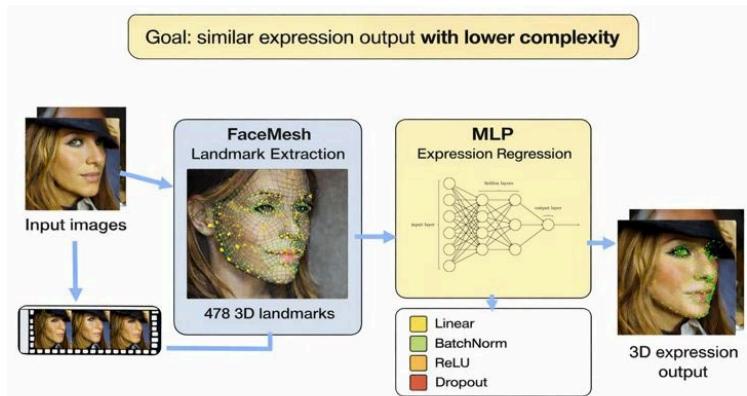
Results dari hasil training dan validation yang kami dapatkan,



Model terbaik : train\_loss=0.047569, val\_loss=0.080277

Pelatihan difokuskan pada jaringan MLP sebagai regression head, sementara Mediapipe facemesh digunakan sebagai landmark extraction yang bersifat fixed.

### 3.3. Diagram Arsitektur:



Gambaran arsitektur model ditunjukkan pada Gambar di atas dengan tujuan menghasilkan output ekspresi wajah 3D yang serupa dengan metode kompleks, namun dengan kompleksitas yang lebih rendah. Proses dimulai dari input citra wajah yang terlebih dahulu diberikan data augmentation berupa in-plane rotation, motion blur, dan out-of-plane rotation untuk meningkatkan variasi data serta meningkatkan robustness model terhadap perubahan pose dan kondisi pencahayaan.

Pada penelitian ini, proses ekstraksi fitur wajah tidak menggunakan backbone CNN secara langsung, melainkan memanfaatkan MediaPipe FaceMesh untuk melakukan ekstraksi landmark wajah 3D. Modul FaceMesh secara internal melakukan face detection sekaligus 3D facial landmark regression yang menghasilkan 478 titik landmark wajah 3D, yang terbukti lebih stabil terhadap variasi pose dan iluminasi.

Landmark wajah yang diperoleh selanjutnya direpresentasikan ke dalam vektor satu dimensi (1D) dan digunakan sebagai input ke dalam Multi Layer Perceptron (MLP). Jaringan MLP berfungsi sebagai expression regression module yang mempelajari hubungan spasial antar titik landmark wajah. Arsitektur MLP terdiri dari beberapa lapisan linear yang dilengkapi dengan Batch Normalization, ReLU activation, dan Dropout untuk meningkatkan stabilitas pelatihan dan mengurangi risiko overfitting.

Output akhir dari jaringan ini adalah representasi ekspresi wajah 3D, yang menggambarkan ekspresi wajah dari citra input secara efisien tanpa memerlukan arsitektur CNN yang kompleks.

#### Landmark Regression Regularization

Pada cabang landmark regression regularization, model memprediksi 136 titik landmark wajah dalam ruang 2D. Loss landmark digunakan sebagai regularisasi tambahan untuk membantu menstabilkan prediksi bentuk wajah serta mencegah terjadinya distorsi geometris pada rekonstruksi wajah.

## Meta-Joint Optimization

Pada bagian meta-joint optimization, model memprediksi parameter 3D Morphable Model (3DMM) yang mencakup pose, shape, dan expression wajah. Parameter ini digunakan untuk merepresentasikan wajah dalam bentuk 3D sehingga ekspresi dan orientasi wajah dapat direkonstruksi secara lebih akurat dan konsisten.

```
class ExpressionHead(nn.Module):
    def __init__(self):
        super().__init__()
        self.net = nn.Sequential(
            nn.Linear(478 * 3, 512),
            nn.BatchNorm1d(512),
            nn.ReLU(inplace=True),
            nn.Dropout(0.2),

            nn.Linear(512, 256),
            nn.BatchNorm1d(256),
            nn.ReLU(inplace=True),
            nn.Dropout(0.2),

            nn.Linear(256, 256),
            nn.BatchNorm1d(256),
            nn.ReLU(inplace=True),

            nn.Linear(256, 128),
            nn.ReLU(inplace=True),

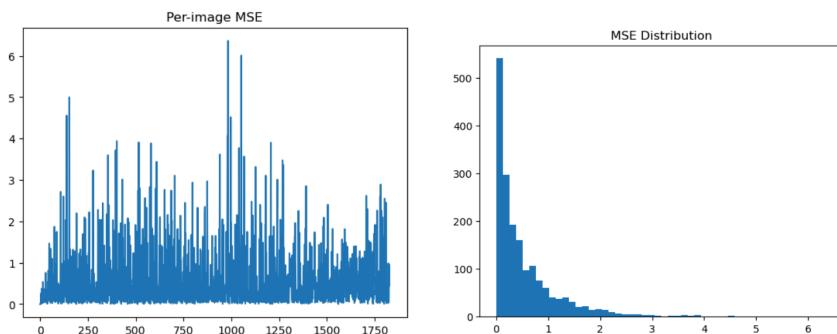
            nn.Linear(128, 10)
        )
```

Ini merupakan gambaran dari expression head untuk MediaPipe FaceMesh + mlp. Setelah melakukan hal tersebut maka proses selanjutnya adalah Multi-Head Output dan Multi-Task Learning namun hal tersebut dibagi menjadi 2 bagian yaitu landmark regression regularization dan meta-join optimization. Pada proses Landmark Regression Regularization itu dia lebih mengarah ke prediksi 136 landmark di wajah, loss landmark digunakan untuk membantu menstabilkan prediction dan mencegah face distortion. Untuk bagian Meta-Join Optimization yang dimana pada bagian ini dia itu lebih ke prediksi Pose, Shape, dan Expression dengan merepresentasikan wajah dalam bentuk 3DMM (3D Morphable Model).

## IV. Evaluation (Evaluasi)

### 4.1. Metrik Evaluasi:

Baseline model 3DDFAv2 (Faceboxes + CNN backbone)



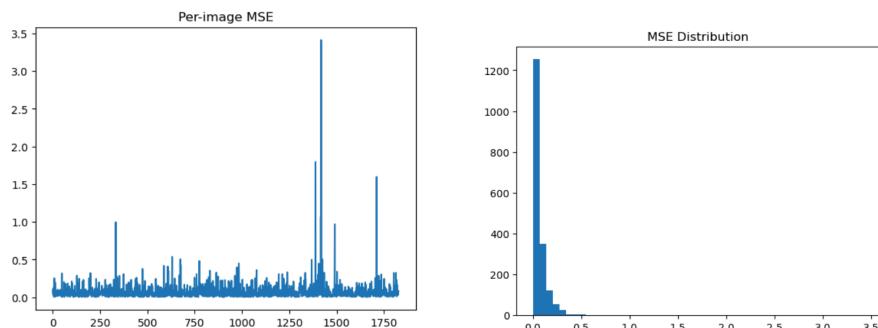
MSE: 0.550077 ,MAE: 0.373496 ,RMSE: 0.741672, R2: -0.972051,  
Mean MSE: 0.55007714, Max MSE: 6.362233

Evaluasi baseline ini menggunakan pipeline 3DDFAv2 dengan Faceboxes sebagai detector, tanpa pelatihan ulang bagian CNN backbone. Model 3DDFAv2 ini dievaluasi dengan menggunakan metrics MSE, MAE, RMSE, dan R2 untuk mengukur akurasi regresi parameter wajah.

Intepretasi metrics:

1. MSE (0.5501): Nilai tinggi, mengindikasikan error kuadrat rata-rata prediksi cukup besar dan adanya prediksi yang menyimpang jauh dari nilai sebenarnya (outlier).
2. MAE (0.3735): Nilai ini mengkonfirmasi error prediksi secara umum yang cukup jauh, bukan hanya disebabkan oleh outlier.
3. RMSE (0.7417): Lebih besar dari MAE, menegaskan distribusi error tidak merata dengan adanya error besar pada sebagian sampel, konsisten dengan Max MSE 6.36 (model gagal regresi stabil pada citra tertentu).
4. R2 (-0.9721): Nilai negatif adalah sinyal kegagalan generalisasi total; performa model lebih buruk daripada sekadar memprediksi nilai rata-rata, tidak mampu menjelaskan variansi data target.

#### MediaPipe FaceMesh + MLP metrics evaluation



MSE: 0.070524, MAE: 0.136519, RMSE: 0.265563, R2: 0.567081, Mean MSE: 0.070523545, Max MSE: 3.4134076

Model usulan ini dievaluasi menggunakan pipeline mediapipe facemesh sebagai modul ekstraksi landmark 3D, diikuti dengan menggunakan MLP sebagai regression head. Evaluasi ini juga menggunakan evaluation metrics yang sama dengan model sebelumnya

Intepretasi Metrik

1. MSE (0.0705): Nilai MSE ini jauh lebih rendah ketimbang hasil dari baseline yang dimana error kuadrat rata-rata berhasil ditekan secara signifikan. Ini menandakan prediksi model lebih dekat dengan groundtruth.
2. MAE (0.1365): Penurunan MAE menunjukkan bahwa secara rata-rata, prediksi terhadap nilai jauh lebih kecil yang memiliki arti peningkatan

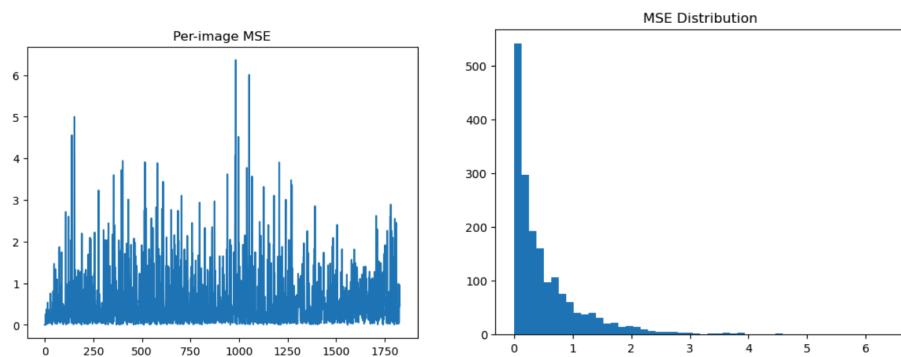
- performance yg tidak hanya disebabkan oleh pengeluaran outlier, tetapi juga oleh peningkatan accuracy.
3. RMSE (0.2656): nilai rmse yang dekat dengan nilai mae mengindikasikan error distribution yang lebih terkendali, tanpa error ekstrem seperti model sebelumnya.
  4. R2 (0.5671): nilai dari r2 ini menunjukkan bahwa model dapat menjelaskan 56% variasi data target. Yang memiliki arti bahwa model memiliki hubungan kuat antara input landmark 3D dan regression parameter berhasil dipelajari oleh model.

#### **4.2. Comparison (Perbandingan):**

Perbandingan antara 2 model yang digunakan antara model baseline 3DDFAv2 (Faceboxes + CNN backbone) dengan model (mediapipe facemesh + MLP). Jika dilihat dari results yang didapatkan model 3DDFAv2 menunjukkan error yang besar, error distribution yang tidak stabil dan R2 negatif yang mengindikasikan kegagalan model dan menjelaskan data. Untuk model kedua kami mendapatkan penurunan signifikan pada MSE, MAE, dan RMSE, serta R2 yang positif yang menandakan regresi yang lebih baik jika dibandingkan dengan model sebelumnya. Peningkatan performance yang didapatkan ini tidak diinterpretasikan sebagai robustness biasa namun dari bukti ini dapat menjelaskan bahwa penggunaan landmark 3D sebagai representasi input downstream dapat menyederhanakan proses regresi dan meningkatkan kualitas dalam memprediksi dibandingkan pendekatan berbasis CNN pada model sebelumnya.

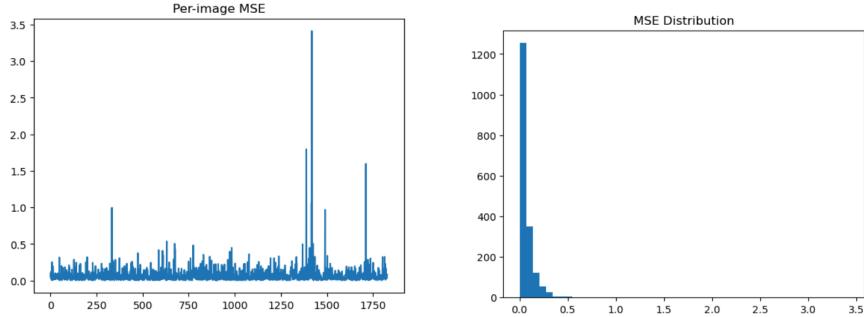
#### **4.3. Visualisasi Hasil:**

Baseline 3DDFAv2 model (Faceboxes + CNN backbone):

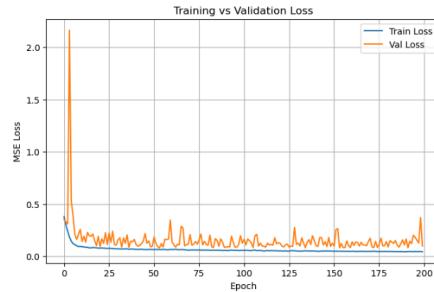


MSE: 0.550077, MAE: 0.373496, RMSE: 0.741672, R2: -0.972051, Mean MSE: 0.55007714, Max MSE: 6.362233

MediaPipe FaceMesh + MLP regression head:



MSE: 0.070524, MAE: 0.136519, RMSE: 0.265563, R2: 0.567081, Mean MSE: 0.070523545, Max MSE: 3.4134076



Model terbaik didapatkan dengan \train\\_loss=0.047569, val\\_loss=0.080277

## V. Deployment (Implementasi)

### 5.1. Aplikasi:

Kami membagi project kami menjadi 3 bagian besar yaitu frontend dan backend dan deployment. Pada bagian frontend kami menggunakan framework next.js + react dengan styling tailwind untuk keseluruhan yang membuka kesempatan untuk menggunakan berbagai UI frameworks seperti shadcn , aceterinity UI dan magic UI serta menggunakan three.js untuk memetakan point plotting pada frontend dalam 3 dimensi. Selain itu frontend kami menggunakan WebRTC untuk mengambil foto dari webcam dan mengirimkan gambar ke backend melalui POST request ke endpoint /infer kami. Pada frontend kami , kami melakukan deployment melalui vercel yang memudahkan deployment secara berkala karena bisa secara otomatis re-deploy pada saat sebuah perubahan terjadi di main branch git repository kami. Pada vercel , kami tidak mengubah banyak selain menambahkan environment variable berupa backend endpoint kami saja.

Pada bagian backend, kami menggunakan fast api dan pytorch untuk menerima REST API yang ringan dan cepat dan memproses hasil gambar yang didapatkan melalui media pipe face mesh untuk mendapatkan 3DMM parameter dan normalisasinya untuk kebutuhan MLP yang digunakan untuk mencari expression parameter dengan batuan torch. Selain itu, kami melakukan deployment backend menggunakan docker container dan azure container apps dimana docker digunakan untuk mengunduh dependensi yang berat seperti pytorch dan media pipe sekaligus memberikan lingkungan virtual yang konsisten sedangkan azure container apps

digunakan untuk mendapatkan endpoint publik tanpa perlu mengatur server secara manual, integrasi dengan azure container registry sebagai tempat penyimpanan sementara dan menyediakan auto scaling untuk deployment berskala besar.

## 5.2. Alur Kerja:

Aplikasi frontend memulai proses dengan melakukan inisialisasi state dan referensi yang digunakan serta mengatur tampilan [three.js](#) seperti posisi kamera, renderer, lighting , dsb. Kemudian saat user menekan tombol start camera maka, web akan memulai webRTC dan mengambil stream sekaligus menangkap gambar statis untuk dikirimkan ke backend endpoint dimana gambar akan diproses pada tahap face detection dan preprocessing menggunakan modul bawaan MediaPipe, yang meliputi pendektsian area wajah, cropping citra, penyesuaian ukuran, serta normalisasi agar sesuai dengan format input yang dibutuhkan oleh model secara internal.

Kemudian pada bagian backend, gambar mentah akan dipreprocess dan dijalankan melalui MediaPipe FaceMesh yang telah dimuat pada saat backend dimulai. Proses pemuatan model ini dilakukan satu kali di awal untuk menghindari beban komputasi berulang dan memastikan efisiensi selama proses inferensi. Model facemesh digunakan secara eksklusif untuk melakukan estimasi landmark wajah tanpa adanya pembaruan parameter.

Kemudian hasil dari facemesh akan diekstraksi menjadi landmarks dan menghasilkan 468 titik landmark wajah dalam koordinat tiga dimensi (x, y, z) yang merepresentasikan struktur geometris wajah secara detail dan konsisten. Kemudian hasil dari landmark tersebut akan digunakan untuk proses inferensi ke model Multilayer Perceptron (MLP) yang berperan sebagai model lanjutan (downstream). MLP digunakan untuk melakukan pemetaan landmark wajah ke 10 padat yang pada kasus ini digunakan untuk klasifikasi ekspresi wajah yaitu mulut terbuka atau tertutup. Kedua hasil tersebut akan dikirim kembali ke frontend dimana [three.js](#) akan memetakan seluruh landmark dan mengambil 2 koefisien untuk mendekripsi mulut terbuka melalui pendekatan threshold yang kami tuning.

## VI. Reflection (Refleksi)

### 6.1. Analisis Kritis:

Proyek ini memodifikasi pipeline 3DDFA\_v2 dengan mengganti FaceBoxes dan CNN menjadi MediaPipe FaceMesh dan model MLP yang lebih ringan. Tujuannya adalah menghasilkan expression parameter yang serupa dengan 3DDFA namun dengan model yang lebih efisien. MediaPipe FaceMesh digunakan untuk mengekstraksi landmark wajah, sementara MLP memetakan landmark tersebut ke expression parameter melalui supervised learning.

Hasil evaluasi menunjukkan bahwa MLP mampu menghasilkan expression parameter yang sangat sesuai dengan ground truth dataset dan bahkan memberikan metrik evaluasi yang lebih baik dibanding pipeline 3DDFA asli. Hal ini menunjukkan bahwa MLP efektif sebagai downstream model karena ringan, tidak over-parameterized, dan tidak memerlukan dataset besar seperti CNN.

Namun, pendekatan ini memiliki keterbatasan, antara lain ketergantungan penuh pada kualitas dataset (noise akan langsung mempengaruhi performa), rendahnya interpretabilitas parameter ekspresi karena tidak direkayasa secara semantik seperti pada 3DDFA, serta ketidakmampuan model mempelajari dependensi temporal karena dataset bersifat statis. Selain itu, kualitas output MLP sangat bergantung pada akurasi landmark yang dihasilkan oleh FaceMesh, terutama pada kondisi wajah terhalangi.

## 6.2. Kendala & Pembelajaran:

Pengembangan sistem konversi wajah 2D ke ekspresi 3D terbukti kompleks dan menuntut pemahaman mendalam terhadap face landmarking, 3DMM, BFM, dan pipeline 3DDFA. Selama eksperimen, ditemukan bahwa MLP dapat mencapai error rendah meskipun tanpa pemaknaan semantik pada setiap dimensi ekspresi, yang memberikan insight bahwa pemetaan numerik dapat efektif namun mengurangi interpretabilitas.

Proyek ini juga memberikan pembelajaran dalam implementasi sistem end-to-end, mulai dari ekstraksi landmark, training model, hingga visualisasi real-time berbasis Three.js, termasuk tantangan sinkronisasi data dan normalisasi koordinat. Dari sisi deployment, keterbatasan layanan cloud free tier mendorong penggunaan Docker dan Azure Container Apps untuk backend, serta Vercel untuk frontend. Terakhir, eksperimen menunjukkan bahwa tidak semua representasi input cocok digunakan untuk mempelajari ekspresi. Beberapa konfigurasi training langsung gagal mempelajari pemetaan karena kompleksitas parameter yang tinggi.

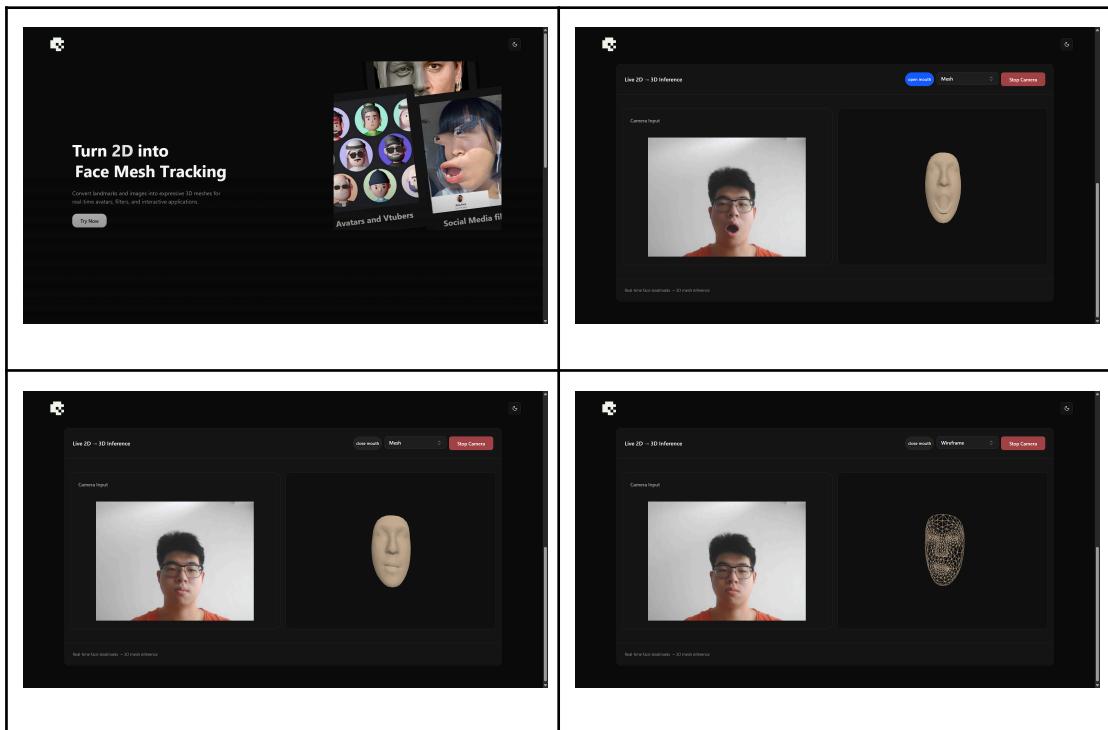
## VII. References (Referensi)

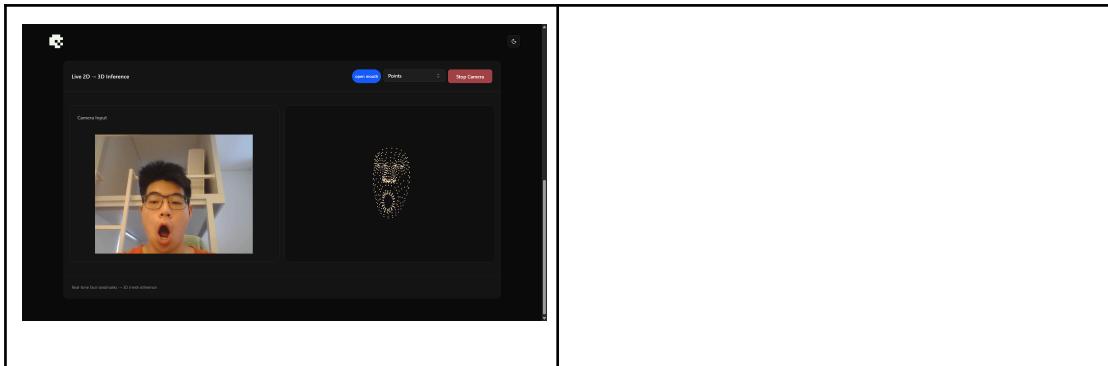
- [1] J. Thies, M. Zollhöfer, M. Nießner, L. Valgaerts, M. Stamminger, and C. Theobalt, “Real-time Expression Transfer for Facial Reenactment.” [Online]. Available: <https://www.nießnerlab.org/papers/2015/11face/thies2015realtime.pdf>
- [2] J. Zhang, K. Chen, and J. Zheng, “Facial Expression Retargeting From Human to Avatar Made Easy,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 2, pp. 1274–1287, Feb. 2022, doi: <https://doi.org/10.1109/tvcg.2020.3013876>.
- [3] X. Xu, Z. Wu, X. Wang, and M. Zhou, “Retargeting 3D facial expressions in real time based on Kinect,” Retargeting 3D Facial Expressions in Real Time Based on Kinect, pp. 209–213, Nov. 2014, doi: <https://doi.org/10.1145/2670473.2670502>.
- [4] Z. Tu, D. Weng, B. Liang, and L. Luo, “3D facial expression retargeting framework based on an identity-independent expression feature vector,” *Multimedia Tools and Applications*, vol. 82, no. 15, pp. 23017–23034, Feb. 2023, doi: <https://doi.org/10.1007/s11042-023-14547-2>.

- [5]T. Costigan, M. Prasad, and R. McDonnell, “Facial retargeting using neural networks,” Proceedings of the Seventh International Conference on Motion in Games, pp. 31–38, Nov. 2014, doi: <https://doi.org/10.1145/2668064.2668099>.
- [6]B. Chaudhuri, “Deep Facial Expression Modeling and 3D Motion Retargeting from 2D Images - ProQuest,” Proquest.com, 2021.  
<https://www.proquest.com/openview/3e05bc6c10c3fdcf61fd6699ed57567c/1?pq-origin=gscholar&cbl=18750&diss=y>
- [7]E. Wood et al., “3D Face Reconstruction with Dense Landmarks,” Lecture notes in computer science, vol. 13673, pp. 160–177, Jan. 2022, doi:  
[https://doi.org/10.1007/978-3-031-19778-9\\_10](https://doi.org/10.1007/978-3-031-19778-9_10).
- [8]J. Wang, J. Zhang, C. Luo, and F. Chen, “Joint head pose and facial landmark regression from depth images,” Computational Visual Media, vol. 3, no. 3, pp. 229–241, May 2017, doi: <https://doi.org/10.1007/s41095-017-0082-8>.
- [9]G. Medioni, J. Choi, M. Labeau, J. T. Leksut, and L. Meng, “3D Facial Landmark Tracking and Facial Expression Recognition,” Journal of information and communication convergence engineering, vol. 11, no. 3, pp. 207–215, Sep. 2013, doi:  
<https://doi.org/10.6109/jicce.2013.11.3.207>.
- [10]Y. Wu and Q. Ji, “Facial Landmark Detection: A Literature Survey,” International Journal of Computer Vision, vol. 127, no. 2, pp. 115–142, May 2018, doi:  
<https://doi.org/10.1007/s11263-018-1097-z>.

## VIII. LAMPIRAN

### *Screenshot Aplikasi*





### Link Repository (*Source Code*)

Github link: <https://github.com/KinsleyTan/Deep-LearningProject>

### Link Deploy

<https://my-deep-learning-project.vercel.app/>

### Link Demo Video

[https://drive.google.com/file/d/1dPx\\_elNgc1Bd4j9QjEE6bLBniwgIFmVj/view?usp=sharing](https://drive.google.com/file/d/1dPx_elNgc1Bd4j9QjEE6bLBniwgIFmVj/view?usp=sharing)

### Link Video Presentation

<https://drive.google.com/file/d/15EgzWn3ckLguXCz9ozwD7KNof26efuMG/view?usp=sharing>

### Link Power Point

[https://www.canva.com/design/DAG7dXxKjsA/3yUudbPPugmvf0Fl2C1hQO/edit?utm\\_content=DAG7dXxKjsA&utm\\_campaign=designshare&utm\\_medium=link2&utm\\_source=sharebutton](https://www.canva.com/design/DAG7dXxKjsA/3yUudbPPugmvf0Fl2C1hQO/edit?utm_content=DAG7dXxKjsA&utm_campaign=designshare&utm_medium=link2&utm_source=sharebutton)

### Link Dataset

<https://www.kaggle.com/datasets/mohamedadlyi/aflw2000-3d>