

MODEL FAST TRANSFER LEARNING PADA JARINGAN SYARAF TIRUAN KONVOLUSIONAL UNTUK KLASIFIKASI GENDER BERDASARKAN CITRA WAJAH

Bambang Krismono Triwijoyo
Universitas Bumigora, Jl. Ismail Marzuki Mataram,
bkrismono@stmikbumigora.ac.id

ABSTRAK

Wajah adalah objek yang menantang untuk dikenali dan dianalisis secara otomatis oleh komputer dalam banyak aplikasi menarik seperti klasifikasi jenis kelamin wajah. Variasi visual yang besar dari objek wajah, seperti oklusi, perubahan pose, dan pencahayaan ekstrem, memberikan tantangan besar untuk tugas-tugas ini dalam aplikasi dunia nyata. Makalah ini menjelaskan model representasi pembelajaran transfer cepat (*Fast Transfer Learning*) melalui penggunaan model jaringan saraf convolutional (CNN) untuk klasifikasi gender dari citra wajah. *Transfer learning* bertujuan untuk menyediakan kerangka kerja untuk memanfaatkan pengetahuan yang diperoleh sebelumnya untuk memecahkan masalah baru tetapi serupa dengan hasil jauh lebih cepat dan efektif. Hasil eksperimen menunjukkan bahwa metode pembelajaran transfer memiliki akurasi lebih cepat dan lebih tinggi daripada jaringan CNN tanpa pembelajaran transfer.

Kata kunci: gender classification; transfer learning; CNN

ABSTRACT

The face is a challenging object to be recognized and analyzed automatically by a computer in many interesting applications such as facial gender classification. The large visual variations of faces, such as occlusions, pose changes, and extreme lightings, impose great challenge for these tasks in real world applications. This paper explained the fast transfer learning representations through use of convolutional neural network (CNN) model for gender classification from face image. Transfer learning aims to provide a framework to utilize previously-acquired knowledge to solve new but similar problems much more quickly and effectively. The experimental results showed that the transfer learning method have faster and higher accuracy than CNN network without transfer learning.

Keywords: gender classification; transfer learning; convolutional neural network

I. PENDAHULUAN

Klasifikasi Gender adalah masalah klasifikasi biner, di mana seseorang harus memprediksi gambar seperti laki-laki atau perempuan. Ini adalah pekerjaan yang mudah bagi manusia, tetapi pekerjaan yang menantang bagi komputer. Sebagian besar model komputasi klasifikasi gender menggunakan gambar seluruh wajah, memberikan bobot yang sama untuk semua area wajah, terlepas dari pentingnya fitur internal [1]. Pemrosesan global sebagian besar berkaitan dengan informasi seperti bentuk dan konfigurasi fitur internal, sementara pemrosesan fitur melibatkan representasi yang lebih terperinci seperti mata, mulut dan bentuk wajah.

Di sisi lain, teknologi mesin pembelajaran telah menarik perhatian yang luar biasa dari para peneliti di berbagai bidang komputasi, khususnya pada pengenalan pola citra, antara lain aplikasi

deteksi citra kendaraan pada sistem transportasi lalu lintas kendaraan dengan tingkat akurasi 95% [2], serta segmentasi citra pembuluh darah retina menggunakan deteksi garis multiskala [3]. Pada sebagian besar teknologi ini bekerja di bawah asumsi umum bahwa data pelatihan dan data uji memiliki ruang fitur yang identik dengan distribusi dari data uji. Akibatnya, setelah ruang fitur atau distribusi fitur dari data uji berubah, model prediksi tidak dapat digunakan dan harus dibangun kembali dan dilatih ulang dari awal menggunakan data pelatihan yang baru dikumpulkan, yang sangat mahal dan kadang-kadang tidak praktis [4]. Demikian pula, karena model berbasis pembelajaran membutuhkan data berlabel yang memadai untuk pelatihan, hampir tidak mungkin untuk membuat model berbasis pembelajaran untuk domain target yang memiliki sedikit data berlabel yang tersedia untuk pembelajaran yang diawasi. Namun, jika kita dapat mentransfer dan mengeksplorasi pengetahuan dari domain sumber

yang sama tetapi tidak identik dengan banyak data berlabel, kita dapat menggunakannya untuk pembangunan model berbasis pembelajaran untuk domain target [4].

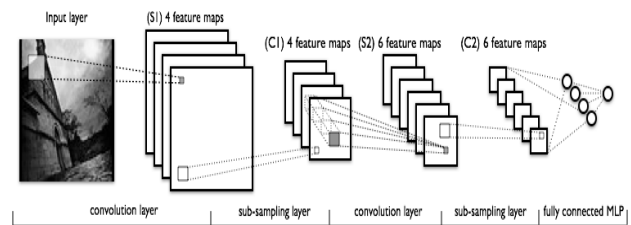
Studi *transfer learning* telah terinspirasi oleh fakta bahwa manusia dapat memanfaatkan pengetahuan yang diperoleh sebelumnya untuk memecahkan masalah baru namun serupa jauh lebih cepat dan efektif [4]. Dalam makalah ini berdasarkan eksperimen kami menunjukkan representasi pembelajaran transfer cepat melalui penggunaan model jaringan saraf convolutional atau *Convolutional Neural Network* (CNN) untuk klasifikasi gender dari citra wajah.

CNN adalah pengembangan dari Multilayer Perceptron (MLP) yang didesain untuk mengolah data dua dimensi. CNN termasuk dalam jenis Deep Neural Network karena kedalaman jaringan yang tinggi dan banyak diaplikasikan pada data citra. Pada kasus klasifikasi citra, MLP kurang sesuai untuk digunakan karena tidak menyimpan informasi spasial dari data citra dan menganggap setiap piksel adalah fitur yang independen sehingga menghasilkan hasil yang kurang baik.

CNN pertama kali dikembangkan dengan nama NeoCognitron oleh Kunihiko Fukushima, seorang peneliti dari NHK Broadcasting Science Research Laboratories, Kinuta, Setagaya, Tokyo, Jepang [5]. Konsep tersebut kemudian dimatangkan oleh Yann LeCun, seorang peneliti dari AT&T Bell Laboratories di Holmdel, New Jersey, USA. Model CNN dengan nama LeNet berhasil diterapkan oleh LeCun pada penelitiannya mengenai pengenalan angka dan tulisan tangan [6]. Pada tahun 2012, Alex Krizhevsky dengan penerapan CNN miliknya berhasil menjuarai kompetisi ImageNet Large Scale Visual Recognition Challenge 2012. Prestasi tersebut menjadi momen pembuktian bahwa metode Deep Learning, khususnya CNN. Metode CNN terbukti berhasil mengungguli metode Machine Learning lainnya seperti SVM pada kasus klasifikasi objek pada citra.

Cara kerja CNN memiliki kesamaan pada MLP, namun dalam CNN setiap neuron dipresentasikan dalam bentuk dua dimensi, tidak seperti MLP yang setiap neuron hanya berukuran satu dimensi. Pada CNN, data yang dipropagasikan pada jaringan adalah data dua dimensi, sehingga operasi linear dan parameter bobot pada CNN berbeda. Pada CNN operasi linear menggunakan operasi konvolusi, sedangkan

bobot tidak lagi satu dimensi saja, namun berbentuk empat dimensi yang merupakan kumpulan kernel konvolusi. CNN terdiri dari berbagai layer dan beberapa neuron pada masing-masing layer. Kedua hal tersebut tidak dapat ditentukan menggunakan aturan yang pasti dan berlaku berbeda-beda pada data yang berbeda [7]. Gambar 1 Mengilustrasikan arsitektur jaringan CNN.



Gambar 1. Arsitektur CNN^[5]

CNN adalah jenis jaringan syaraf tiruan feed-forward dimana pola konektivitas antara neuronnya terinspirasi oleh pengorganisasian korteks visual pada jaringan syaraf biologis. Neuron korteks individu merespons rangsangan di wilayah terbatas yang dikenal sebagai bidang reseptif. Bidang reseptif neuron yang berbeda sebagian tumpang tindih sehingga mereka menghiasi bidang visual.

Respon neuron individu terhadap rangsangan di dalam medan reseptifnya dapat didekati secara matematis dengan operasi konvolusi. Jaringan konvolusi terinspirasi oleh proses biologis dan merupakan variasi perceptron multilayer yang dirancang untuk menggunakan jumlah preprocessing minimal. CNN memiliki aplikasi yang luas seperti: pengenalan gambar dan video, sistem recommender dan pemrosesan bahasa alami.

CNN menggunakan tensor orde 3, dimana sebuah citra masukan berukuran $H \times W$ piksel dengan 3 kanal (kanal warna R, G dan B) diproses melalui serangkaian tahapan proses secara berurutan. Satu tahapan proses disebut lapisan (layer) yang terdiri dari lapisan konvolusional (convolutional layer), lapisan pooling (pooling layer), lapisan normalisasi (normalization layer), lapisan terhubung penuh (fully connected layer). Diskripsi abstrak dari struktur CNN data di ekspresikan sebagai berikut [6]:

$$x^1 \rightarrow w^1 \rightarrow x^2 \rightarrow \dots \rightarrow x^{L-1} \rightarrow x^L \rightarrow w^L \rightarrow z \quad (1)$$

CNN beroperasi secara berurutan lapisan per lapisan dimana x^1 sebagai masukan berupa citra atau *image* dalam bentuk tensor orde 3 yang akan diproses pada lapisan pertama yaitu w^1 yang berupa tensor dan menghasilkan keluaran x^2 , dimana x^2 sekaligus merupakan masukan untuk lapisan kedua, demikian seterusnya sampai pada keluaran x^L . Sehingga pada lapisan ke- l dengan masukan berupa tensor orde 3 yaitu x^l dengan $x^l \in \mathbb{R}^{H^l \times W^l \times H^l}$, sehingga diperlukan *triple index set* (i^l, j^l, d^l) untuk mengacu ke satu elemen x^l , pada kanal ke d^l dan pada lokasi spasial (i^l, j^l) yaitu pada baris ke i^l dan kolom ke j^l .

Pada proses pembelajaran CNN digunakan metode strategi *mini-batch* sehingga x^l menjadi tensor orde 4 dengan $x^l \in \mathbb{R}^{H^l \times W^l \times H^l \times N^l}$, dimana N adalah ukuran *mini-batch*.

Pada lapisan terakhir ditambahkan metode pembelajaran propagasi balik (*back propagation*) pada lapisan *fully connected* z , dimana pada permasalahan klasifikasi citra menjadi sejumlah C kelas, maka output x^L dihubungkan dengan vektor dengan C dimensi.

Penelitian tentang klasifikasi gender otomatis kembali ke awal 1990-an. Hasil pertama dilaporkan secara bersamaan oleh [8][9]. Dengan cara yang sama seperti dengan deteksi wajah, metode klasifikasi gender dapat secara kasar dibagi dalam metode berbasis fitur dan penampilan. Dua metode pertama berbasis penampilan dan keduanya menggunakan pendekatan jaringan saraf multi-layer. Wajah disejajarkan secara manual untuk eksperimen. Banyak metode yang berbeda telah dicoba setelah dua metode pertama diterbitkan.

Model grafik ditempatkan secara manual ke wajah dan klasifikasi gender didasarkan pada *wavelet Gabor* yang ditempatkan pada node model yang diusulkan oleh [10]. Mereka juga menggunakan sistem untuk pengenalan wajah. Bereksperimen dengan gambar wajah beresolusi sangat rendah dan jaringan saraf telah diusulkan oleh [11]. Mereka mencapai tingkat klasifikasi 93% dengan hanya gambar wajah ukuran 8x8 piksel. *Gabor wavelet* dengan analisis komponen utama atau *principal component analysis* (PCA) dan analisis diskriminan linier atau *linear discriminant analysis* (LDA) digunakan oleh [12]

untuk mendeteksi wajah dan mengklasifikasikan gender. Grafik yang mirip dengan yang ditempatkan pada wajah secara otomatis seperti yang telah diterapkan oleh [10]. Mereka mencapai tingkat klasifikasi 92% yang sedikit lebih tinggi dari tingkat klasifikasi 91,3% yang dicapai oleh [10].

Pemilihan fitur adalah masalah penting untuk klasifikasi gender dan menunjukkan bahwa algoritma genetika atau *genetic algorithms* (GA) cocok untuk tugas tersebut [13]. Mereka membuat vektor fitur dari gambar wajah menggunakan PCA. Kemudian mereka memilih subset fitur dari vektor menggunakan algoritma genetika dan menggunakan fitur sebagai input ke pengelompokan gender. Kinerja empat pengklasifikasi berbeda dibandingkan: Bayesian, jaringan saraf tiruan, *Support Vector Machines* (SVM) dan analisis diskriminan linier. Klasifikasi SVM mencapai tingkat klasifikasi terbaik 95,3%.

Analisis komponen independen atau *Independent Component Analysis* (ICA) telah digunakan oleh [14], untuk mengekstrak fitur dari gambar wajah dan LDA untuk mengklasifikasikan gender. Mereka mencapai tingkat klasifikasi 99,3% yang mengesankan dengan gambar wajah FERET yang dipotong dan dinormalisasi secara manual. SVM jarang telah digunakan oleh [15] untuk mengklasifikasikan gender, dengan akurasi tingkat klasifikasi 94,42%.

Dua pengklasifikasi yang berbeda: *Self Organizing Map* (SOM) dan *threshold Adaboost* telah dicoba oleh [16]. Kebaruan dengan pendekatan mereka adalah bahwa mereka menggunakan Pola Biner Lokal atau *Local Binary Patterns* (LBP) untuk membuat fitur untuk input. Tingkat klasifikasi terbaik, 95,75%, dicapai dengan *classifier Adaboost*. [17] Juga bereksperimen dengan LBP namun, mereka menggunakan *Support Vector Machine* sebagai *classifier* dan mencapai tingkat klasifikasi 96,75%.

Operator perbandingan piksel dengan *Adaboost classifier* telah digunakan oleh [18] dan mencapai lebih dari 93% akurasi klasifikasi yang melampaui akurasi *classifier SVM* yang menggunakan piksel sebagai input. Fitur yang mereka gunakan adalah pilihan yang menarik untuk sistem waktu-nyata yang menggabungkan deteksi wajah dan klasifikasi gender karena operator nilai perbandingan dapat dihitung dengan cepat. Mereka mencapai klasifikasi 50 kali lebih cepat dengan *Adaboost classifier* daripada dengan SVM

dan juga melakukan analisis sensitivitas untuk pengklasifikasi dengan melakukan variasi rotasi, skala, dan offset yang diimplementasikan dalam bidang gambar wajah. Jenis ketidakakuratan ini kemungkinan terjadi ketika wajah terdeteksi dan juga memengaruhi akurasi klasifikasi gender.

Terinspirasi oleh kinerja yang baik dari CNN dalam tugas-tugas visi komputer, seperti klasifikasi gambar [19] dan pengenalan wajah [20], beberapa pendekatan klasifikasi gender berbasis CNN telah diusulkan. CNN untuk klasifikasi gender yang kuat dalam lingkungan yang tidak dibatasi telah diusulkan oleh [21]. Mereka menguji metode mereka pada basis data LFWA [22] dan mendapatkan kinerja terbaik 98,8% untuk klasifikasi gender. Dibandingkan dengan klasifikasi objek menggunakan dataset skala besar, Imagenet [23]. Hanya ada beberapa dataset skala kecil yang dikumpulkan dari dunia nyata untuk klasifikasi gender dan senyuman. Pendekatan berbasis CNN yang bekerja dengan baik pada dataset kecil telah diusulkan oleh [24].

Model CNN yang mendalam juga telah digunakan untuk klasifikasi gender dalam beberapa tahun terakhir, model GiNet yang diusulkan dalam [24], dilatih dengan dataset Adience [25] dari awal untuk estimasi gender. Dalam [26], model ensemble CNN diusulkan untuk berjalan secara efisien di perangkat yang disematkan. Eksperimen metode ini dilakukan pada dataset LFW [27] dan hasil mutakhir diperoleh sesuai dengan [26]. Studi di [28] menggabungkan fitur lokal dan jaringan dalam dengan belajar dari tumpang tindih lokal tambalan dengan cara umpan maju.

Penelitian tentang *transfer learning* telah dilakukan sejak 1995 dengan berbagai nama: belajar untuk belajar; belajar sepanjang hayat; transfer pengetahuan; pembelajaran meta; transfer induktif; konsolidasi pengetahuan; pembelajaran sensitif konteks dan pembelajaran multi-tugas [29]. Domain, yang dilambangkan dengan

$$D=\{X, P(X)\} \quad (2)$$

Persamaan (1) terdiri dari dua komponen yaitu ruang fitur X dan distribusi probabilitas marginal $P(X)$, di mana $X=\{x_1, \dots, x_n\} \in X$. Suatu tugas, yang dilambangkan dengan

$$T=\{Y, f(\cdot)\} \quad (3)$$

Persamaan (2) terdiri dari dua komponen yang merupakan ruang label $Y=\{y_1, \dots, y_m\}$ dan Fungsi prediksi yang objektif $f(\cdot)$ yang tidak diamati dan

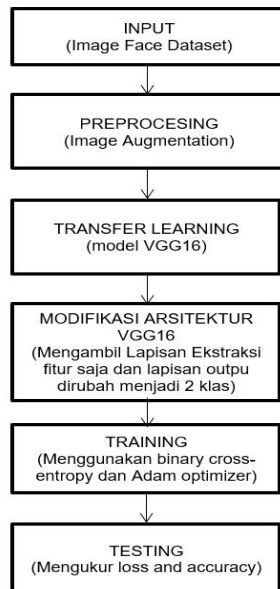
harus dipelajari secara berpasangan $\{x_i, y_i\}$. Diberikan domain sumber D_s dan *learning task* T_s , target domain D_t dan target learning task T_t , transfer learning bertujuan untuk meningkatkan pembelajaran fungsi prediksi target $f_t(\cdot)$ pada domain D_t menggunakan pengetahuan pada D_s dan T_s dimana $D_s \neq D_t$ or $T_s \neq T_t$.

Arsitektur multi-tahap yang terdiri dari lapisan bergantian dari konvolusi dan max pooling untuk mengekstraksi fitur data telah diusulkan oleh [30]. Model baru yang memadukan struktur di atas dan beberapa tugas diusulkan untuk pembelajaran transfer [31]. Dalam model ini, tugas target dan tugas terkait dilatih bersama dengan input bersama dan lapisan tersembunyi, dan secara terpisah menghasilkan neuron. Model ini kemudian diperluas ke kasus di mana setiap tugas memiliki beberapa neuron keluaran [32].

II. METODOLOGI

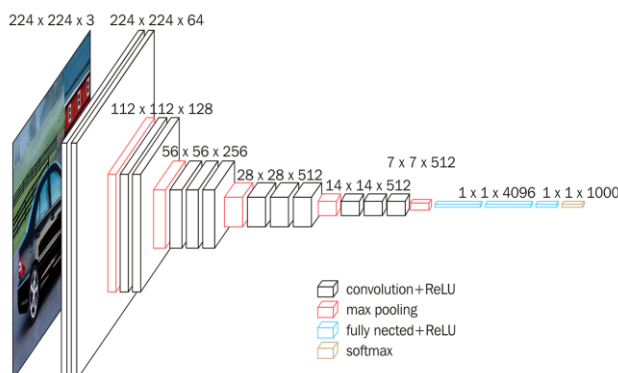
Metode yang digunakan untuk klasifikasi gender ditunjukkan pada Gambar 2. Langkah pertama adalah preprocessing untuk menambah data menggunakan generator data gambar. Setelah preprocessing, proses selanjutnya adalah transfer learning menggunakan VGG16 dari ImageNet [19]. Langkah selanjutnya adalah modifikasi arsitektur model VGG16 sesuai dengan masalah klasifikasi gender, dimana hanya menggunakan lapisan ekstraksi fitur dari VGG16 serta merubah lapisan output menjadi hanya 2 klas.

Tahap selanjutnya adalah pelatihan, model klasifikasi 2 kelas atau yang biasa disebut klasifikasi biner, di mana 0 untuk kelas pria dan 1 untuk kelas wanita. Kami menggunakan Sigmoid sebagai fungsi aktivasi di lapisan output dan untuk semua lapisan tersembunyi kami menggunakan *rectified linear unit* (ReLU) dan menggunakan cross-entropy biner dan fungsi Adam untuk optimizer. Tahap terakhir adalah pengujian untuk mengukur keakuratan hasil.



Gambar 2. Metode Klasifikasi Gender.

Gambar 3 adalah arsitektur VGG16. Pada penelitian ini hanya menggunakan lapisan Ekstraksi Fitur saja, dan menyimpan hasil pelatihan ke file. Lapisan *Full Connected* di VGG16 terdiri dari 4.096 neuron di lapisan tersembunyi dan 1000 neuron di lapisan keluaran, karena ImageNet memiliki 1000 kelas. Sementara dataset citra gender hanya memiliki 2 kelas (Pria/Wanita), jadi layer output direvisi menjadi 2 kelas.



Gambar 3. Arsitektur VGG16 [16].

A. Dataset

Dataset yang digunakan dikumpulkan dari *UI Faces API* dan *Random User Generator* yang dapat diunduh di <https://uifaces.co/api-docs> ; <https://randomuser.me/photos> Ada 800 foto untuk data pelatihan dan 240 foto untuk pengujian yang dibagi menjadi 2 kelas yaitu pria dan wanita.

Sampel dataset seperti yang ditunjukkan pada Gambar 4, gambar dari *Random User Generator* memiliki ukuran yang sama 224 x 224 piksel tetapi foto dari Wajah UI memiliki ukuran yang berbeda. Jadi kita harus melakukan pra-proses terlebih dahulu untuk mendapatkan ukuran foto yang seragam sehingga bisa lebih mudah saat melakukan pelatihan.



Gambar 4. Sampel dataset. (Sumber: *UI Faces API*)

B. Preprocessing

Dari dataset terdapat 400 foto pria dan 400 foto wanita, jumlah data masih kurang memadai untuk mendapatkan kinerja yang optimal. Jadi perlu untuk menambah data. Augmentasi Data adalah teknik untuk memanipulasi data tanpa kehilangan inti atau esensi data [19]. Dalam percobaan kami, kami menggeser, memperbesar, dan membalik sedangkan parameter skala ulang yang kami gunakan membagi nilai Merah Hijau Biru (RGB) dari 0-255 dengan 255, jadi kami mendapatkan nilai RGB dalam kisaran 0-1. Untuk pengujian data, kami hanya melakukan skala ulang. Selanjutnya, *Image Data Generator* digunakan untuk mengubah data dalam bentuk gambar mentah menjadi dataset hasil augmentasi yang digunakan untuk pelatihan dan pengujian.

C. Model Arsitektur

Kami akan menggunakan model VGG-16 [19] yang telah dilatih dalam data ImageNet dengan menciptakan arsitektur yang identik dengan VGG-16 tetapi tanpa lapisan *full connected* dan mengurangi bobotnya. Keras menyediakan beberapa model ImageNet yang populer. Semua bobot VGG-16 telah dilatih menggunakan dataset ImageNet dan dapat mengenali warna, tekstur, dll.

Jadi bobot VGG-16 digunakan untuk mengekstrak fitur dari semua foto dalam dataset.

Tabel 1. Arsitektur Model

Layer	Type	Neurons & Maps	Size	Padding	Stride
0	Input	224×224×3	-	-	-
1	Conv.2D	224×224×64	3×3	1	1
2	Conv.2D	224×224×64	3×3	1	1
3	MaxPool. 2D	112×112×128	3×3	1	2
4	Conv.2D	112×112×128	3×3	1	1
5	Conv.2D	112×112×128	3×3	1	1
6	MaxPool. 2D	56×56×256	3×3	1	2
7	Conv.2D	56×56×256	3×3	1	1
8	Conv.2D	56×56×256	3×3	1	1
9	Conv.2D	56×56×256	3×3	1	1
10	MaxPool. 2D	28×28×512	3×3	1	2
11	Conv.2D	28×28×512	3×3	1	1
12	Conv.2D	28×28×512	3×3	1	1
13	Conv.2D	28×28×512	3×3	1	1
14	MaxPool. 2D	14×14×512	3×3	1	2
15	Conv.2D	14×14×512	3×3	1	1
16	Conv.2D	14×14×512	3×3	1	1
17	Conv.2D	14×14×512	3×3	1	1
18	MaxPool. 2D	7×7×512	3×3	1	2
19	Full Connected	32	1×1	-	-
20	Output	1	1×1	-	-

Seperti yang ditunjukkan pada Tabel 1, kami menggunakan lapisan *Full Connected* (FC) yang merupakan 32 neuron pada layer tersembunyi dan 1 neuron pada lapisan output dengan aktivasi sigmoid.

III. HASIL DAN PEMBAHASAN

Kami menggunakan VGG-16 menggunakan aplikasi paket dari Keras tanpa harus mengaturnya lapis demi lapis dan mengunduh bobotnya. Kami tidak menggunakan lapisan FC dari VGG-16.

Sehingga ketika proses prediksi dilakukan untuk model ini, apa yang akan terjadi adalah bahwa dataset akan mengalir ke lapisan ekstraksi fitur VGG-16. Hasilnya adalah peta fitur dan disimpan dalam file fitur pelatihan dan file fitur validasi yang digunakan untuk melakukan pelatihan pada lapisan FC pada model. Selanjutnya, peta fitur digunakan sebagai data pengujian untuk lapisan FC.

Fungsi Sigmoid digunakan sebagai fungsi aktivasi di lapisan output dan fungsi aktivasi ReLU untuk semua lapisan tersembunyi atau *hidden layer*. Fungsi *cross-entropy* biner digunakan pada pelatihan dan ujicoba model dan fungsi optimasi Adam untuk Mengoptimalkan hasil. Model pelatihan dilakukan dengan menggunakan *stochastic gradient decent* (SGD) dengan ukuran batch gambar enam puluh gambar, 100 *epochs* dan tingkat pembelajaran awal atau *learning rate* adalah 0,0001. Berikut adalah implementasi kode Matlab dari model pelatihan klasifikasi gender menggunakan *Fast transfer learning* VGG-16.

```
# Dimensi Citra
img_width, img_height = 224, 224
```

```
train_data_dir = 'dataset/train'
validation_data_dir =
'dataset/validation'
nb_train_samples = 800
nb_validation_samples = 240
epochs = 100
batch_size = 16
```

```
# Inisialisasi model VGG16
model = applications.VGG16
include_top=False, weights='imagenet')
```

```
# Augmentasi data pelatihan
train_datagen = ImageDataGenerator(
    rescale=1. / 255,
    shear_range=0.2,
    zoom_range=0.2,
    horizontal_flip=True)
```

```
# Generator untuk data pelatihan
train_generator =
train_datagen.flow_from_directory(
    train_data_dir,
    target_size=(img_width, img_height),
    batch_size=batch_size,
    class_mode='binary')
train_features = model.predict_generator(
    train_generator, nb_train_samples //
    batch_size, verbose=1)
np.save('train_features.npy',
    train_features)
```

```
# Generator untuk data pengujian
validation_generator =
test_datagen.flow_from_directory(
```

```

validation_data_dir,
target_size=(img_width, img_height),
batch_size=batch_size,
class_mode='binary')
validation_features =
model.predict_generator(
validation_generator,
nb_validation_samples // batch_size,
verbose=1)
np.save('val_features.npy',
validation_features)

# Inisialisasi FC Layer
fc_model = Sequential()
fc_model.add(Flatten(input_shape=train_data.
shape[1:]))
fc_model.add(Dense(32,
activation='relu'))
fc_model.add(Dense(1,
activation='sigmoid'))

# Fungsi Optimasi Adam dan Cross Entropy
adam = Adam(lr=0.0001)
fc_model.compile(optimizer=adam,
loss='binary_crossentropy',
metrics=['accuracy'])

fc_model.fit(train_data, train_labels,
epochs=epochs,
batch_size=batch_size,

```

```

validation_data=(validation_data,
validation_labels)),

fc_model.save_weights('vgg16.h5')

```

Sedangkan berikut adalah implementasi model pengujian klasifikasi gender menggunakan transfer learning VGG-16.

```

img_width, img_height = 224, 224

test_data_dir = 'dataset/testing'

def get_features():
    datagen =
ImageDataGenerator(rescale=1. / 255)

    # build the VGG16 network
    model =
applications.VGG16(include_top=False,
weights='imagenet')

    generator =
datagen.flow_from_directory(
test_data_dir,
target_size=(img_width,
img_height),
batch_size=1,
class_mode='binary',
shuffle=False)
    features = model.predict_generator(
generator, 8)
    return features

def get_score():
    features = get_features()

```

```

model = Sequential()

model.add(Flatten(input_shape=features.sh
ape[1:]))
model.add(Dense(32,
activation='relu'))
model.add(Dense(1,
activation='sigmoid'))

model.load_weights('vgg16.h5')

score = model.predict(features)
print(score)

get_score()

```

Model tersebut kemudian dibandingkan dengan model CNN lainnya tanpa transfer learning dan dilatih dengan arsitektur seperti yang ditunjukkan pada tabel 2. Kemudian model diuji menggunakan dataset yang sama.

Tabel 2. Model Arsitektur tanpa Transfer learning

Layer	Type	Neurons & Maps	Size	Padding	Stride
0	Input	224×224×3	-	-	-
1	Conv.2D	74×74×16	5×5	1	3
2	Conv.2D	24×24×32	5×5	-	3
3	MaxPool.2D	12×12×32	2×2	-	2
4	Conv.2D	10×10×64	3×3	-	1
5	Conv.2D	8×8×64	3×3	-	1
7	Full Conected	32	-	-	-
8	Output	1	-	-	-

Ketiga saluran warna dari citra wajah diproses secara langsung oleh jaringan. Citra masukan diubah ukurannya menjadi 244×244 piksel. 16 filter ukuran 5×5 piksel diterapkan pada input di lapisan konvolusional pertama, diikuti oleh operator aktivasi linier ReLU, dengan *stride* atau jarak langkah per tiga piksel. Output yang berdimensi 74 × 74 × 16 dari lapisan sebelumnya kemudian diproses oleh lapisan konvolusional kedua, berisi 32 filter ukuran 5 × 5 piksel. Sekali lagi, ini diikuti oleh ReLU, lapisan *max pooling* mengambil nilai maksimal 2 × 2 daerah dengan *stride* atau langkah dua piksel sebagai lapisan *local response normalization*.

Lapisan konvolusional ketiga beroperasi pada ukuran 10 × 10 dengan menerapkan 64 filter ukuran 3 × 3 piksel, diikuti oleh ReLU. Lapisan konvolusional keempat beroperasi pada 8 × 8

dengan menerapkan satu set 64 filter ukuran 3×3 piksel.

Lapisan *full connected* menerima output dari lapisan konvolusional keempat yang berisi 32 neuron diikuti oleh lapisan ReLU. Akhirnya, output dari lapisan *full connected* yang terakhir diumpungkan ke lapisan soft-max berisi satu neuron untuk dua kelas. Berikut adalah implementasi model pelatihan klasifikasi gender tanpa menggunakan transfer learning.

```
# Dimensi Citra
img_width, img_height = 224, 224

train_data_dir = 'dataset/train'
validation_data_dir =
    'dataset/validation'
nb_train_samples = 800
nb_validation_samples = 240
epochs = 100
batch_size = 16

# Augmentasi data pelatihan
train_datagen = ImageDataGenerator(
    rescale=1. / 255,
    shear_range=0.2,
    zoom_range=0.2,
    horizontal_flip=True)

# Generator untuk data pelatihan
train_generator =
    train_datagen.flow_from_directory(
        train_data_dir,
        target_size=(img_width, img_height),
        batch_size=batch_size,
        class_mode='binary')

# Generator untuk data pengujian
validation_generator =
    test_datagen.flow_from_directory(
        validation_data_dir,
        target_size=(img_width, img_height),
        batch_size=batch_size,
        class_mode='binary')

# Inisialisasi Lapisan ekstraksi fitur
inputs = Input(shape=(img_width,
    img_height, 3))
conv_layer = Conv2D(16, (5, 5),
    strides=(3,3),
    activation='relu')(inputs)
conv_layer =
    ZeroPadding2D(padding=(1,1))(conv_layer)
conv_layer = Conv2D(32, (5, 5),
    strides=(3,3),
    activation='relu')(conv_layer)
conv_layer = MaxPooling2D((2,
    2))(conv_layer)
conv_layer = Conv2D(64, (3, 3),
    strides=(1,1),
    activation='relu')(conv_layer)
conv_layer = Conv2D(64, (3, 3),
    strides=(1,1),
    activation='relu')(conv_layer)
```

```
# Lapisan Flatten
flatten = Flatten()(conv_layer)

# Lapisan Fully Connected
fc_layer = Dense(32,
    activation='relu')(flatten)
outputs = Dense(1,
    activation='sigmoid')(fc_layer)

model = Model(inputs=inputs,
    outputs=outputs)

# Inisialisasi Fungsi optimasi Adam dan
    Cross Entropy
adam = Adam(lr=0.0001)
model.compile(optimizer=adam,
    loss='binary_crossentropy',
    metrics=['accuracy'])

model.fit_generator(
    train_generator,
    steps_per_epoch=nb_train_samples //
        batch_size,
    epochs=epochs,
    validation_data=validation_generator,
    validation_steps=nb_validation_samples
        // batch_size,
    callbacks=[callbacks])

model.save_weights('kornet.h5')
```

Dan berikut adalah implementasi model pengujian klasifikasi gender tanpa menggunakan transfer learning

```
img_width, img_height = 224, 224

test_data_dir = 'dataset/testing'

def get_features():
    datagen =
        ImageDataGenerator(rescale=1. / 255)

    # build the VGG16 network
    model =
        applications.VGG16(include_top=False,
            weights='imagenet')

    generator =
        datagen.flow_from_directory(
            test_data_dir,
            target_size=(img_width,
                img_height),
            batch_size=1,
            class_mode='binary',
            shuffle=False)
    features = model.predict_generator(
        generator, 8)
    return features

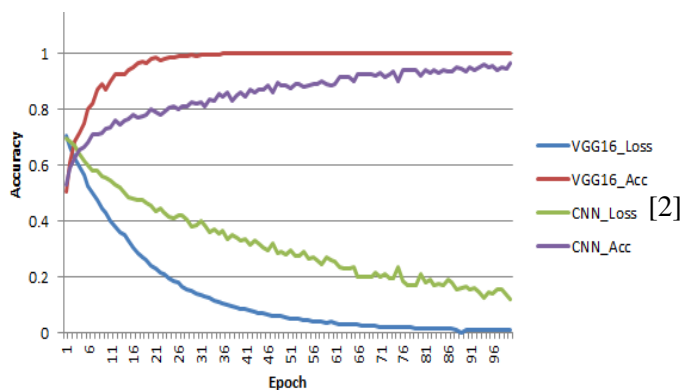
def get_score():
    features = get_features()
    model = Sequential()
```



```
model.add(Flatten(input_shape=features.shape[1:]))
model.add(Dense(32,
activation='relu'))
model.add(Dense(1,
activation='sigmoid'))
model.load_weights('vgg16.h5')
score = model.predict(features)
print(score)

get_score()
```

Model tersebut diimplementasikan menggunakan Python. Pelatihan dilakukan pada Komputer Intel Core i7 Processor, memori 12 GB dan mesin GPU NVIDIA Geforce 940MX. Gambar 4 menyajikan hasil perbandingan untuk klasifikasi gender masing-masing model arsitektur, kami mengukur dan membandingkan akurasi pelatihan dua model dengan dan tanpa transfer learning. Jaringan pelatihan tanpa pembelajaran transfer memiliki akurasi 98% dan diperlukan sekitar 1,3 jam, 961,43 mikro detik per langkah rata-rata. Dan model pelatihan dengan transfer learning memiliki akurasi 100% dan hanya membutuhkan 102 detik dan 1,1 mikro detik per langkah.



Gambar 5. Training performance.

Dari Gambar 5, dapat dilihat bahwa kinerja *loss* akurasi validasi pada model yang menggunakan pembelajaran transfer VGG16 adalah 0,0022 dan *loss* akurasi validasi model tanpa CNN tanpa transfer belajar adalah 0,0654, sehingga overfitting proses validasi dalam model menggunakan transfer learning lebih baik dari model tanpa transfer learning.

IV. SIMPULAN DAN SARAN

Metode klasifikasi gender berdasarkan citra wajah menggunakan *fast transfer learning* dapat mempercepat proses pelatihan dan menambah akurasi hasil. Penelitian ini menggunakan hasil pembelajaran VGG16 dari Imagenet dengan modifikasi arsitektur jaringan di lapisan *full connected* dan lapisan keluaran, serta menggunakan lapisan fitur ekstraksi VGG16. Hasil percobaan menunjukkan bahwa metode pembelajaran transfer memperoleh akurasi lebih cepat dan lebih tinggi dibandingkan dengan jaringan CNN tanpa pembelajaran transfer. Masih ada beberapa aspek yang layak untuk ditingkatkan untuk penelitian lanjutan, antara lain yaitu overfitting yang dapat diatasi dengan menggunakan *Full Connected Layer* yang lebih sederhana, menggunakan *Dropout*, melakukan augmentasi yang lebih agresif atau menambahkan data pelatihan.

REFERENSI

- [1] Buchala, S., Davey, N., Frank, R. J., Gale, T. M., Loomes, M. J., and Kanargard, W. "Gender classification of face images: The role of global and feature-based information". In International Conference on Neural Information Processing. pp. 763-768, 2004. https://doi.org/10.1007/978-3-540-30499-9_117
- [2] Budiarto, J., & Qudsi, J. "Deteksi Citra Kendaraan Berbasis Web Menggunakan Javascript Framework Library". Matrik: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer, vol.18 no.1, pp.125-133, 2018. <https://doi.org/10.30812/matrik.v18i1.325>
- [3] Triwijoyo, B. K. "Segmentasi Citra Pembuluh Darah Retina Menggunakan Metode Deteksi Garis Multi Skala". Matrik: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer, vol. 15, pp.1, pp.13-19, 2015. <https://doi.org/10.30812/matrik.v15i1.28>
- [4] Lu, J., Behbood, V., Hao, P., Zuo, H., Xue, S., and Zhang, G. "Transfer learning using computational intelligence: a survey". Knowledge-Based Systems, no. 80, pp. 14-23, 2015. <https://doi.org/10.1016/j.knosys.2015.01.010>

- [5] Fukushima K and Miyake S. "Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition". In *Competition and cooperation in neural nets* pp. 267-285, 1982. https://doi.org/10.1007/978-3-642-46466-9_18
- [6] LeCun Y, Boser B, Denker J.S, Henderson D, Howard R.E, Hubbard W, and Jackel L.D. "Backpropagation applied to handwritten zip code recognition. *Neural computation*", 1989 <https://doi.org/10.1162/neco.1989.1.4.541>
- [7] Stathakis D. "How many hidden layers and nodes?". *International Journal of Remote Sensing*, vol. 30, no.8, pp. 2133-2147, 2009 <https://doi.org/10.1080/01431160802549278>
- [8] Cottrell, G.W., Metcalfe, J. "EMPATH: Face, emotion, and gender recognition using holons". In: Lippmann, R., Moody, J.E., Touretzky, D.S. (Eds.), *Proc. Advances in Neural Information Processing Systems 3 (NIPS)*. Morgan Kaufmann, pp. 564–571, 1990 <https://dl.acm.org/citation.cfm?id=2986843>
- [9] Golomb, B.A., Lawrence, D.T., Sejnowski, T.J. "SEXNET: A neural network identifies sex from human faces". In: Lippmann, R., Moody, J.E., Touretzky, D.S. (Eds.), *Proc. Advances in Neural Information Processing Systems 3 (NIPS)*. Morgan Kaufmann, pp. 572–579, 1990. <https://dl.acm.org/citation.cfm?id=2986844>
- [10] Wiskott, L., Fellous, J.-M., Krüger, N., von der Malsburg, C., "Face recognition by elastic bunch graph matching". In: Sommer, G., Daniilidis, K., Pauli, J. (Eds.), 7th *International Conference on Computer Analysis of Images and Patterns, CAIP'97*, Kiel. Springer-Verlag, Heidelberg, pp. 456–463, 1997. https://doi.org/10.1007/3-540-63460-6_150
- [11] Tamura, S., Kawai, H., Mitsumoto, H. "Male/female identification from 8 to 6 very low resolution face images by neural network". *Pattern Recognition*, vol. 29, no.2, pp. 331–335, 1996. [https://doi.org/10.1016/0031-3203\(95\)00073-9](https://doi.org/10.1016/0031-3203(95)00073-9)
- [12] Lyons, M., Budynek, J., Plante, A., Akamatsu, S. "Classifying facial attributes using a 2-d Gabor wavelet representation and discriminant analysis". In: *Proc. Internat. Conf. on Automatic Face and Gesture Recognition (FG'00)*, IEEE, Grenoble, France, pp. 202–207, 2000. <https://doi.org/10.1109/AFGR.2000.840635>
- [13] Sun, Z., Bebis, G., Yuan, X., Louis, S.J., December. "Genetic feature subset selection for gender classification: A comparison study". In: *Proc. IEEE Workshop on Applications of Computer Vision (WACV'02)*, pp. 165–170, 2002. <https://doi.org/10.1109/ACV.2002.1182176>
- [14] Jain, A., Huang, J. "Integrating independent components and linear discriminant analysis for gender classification". In: *Proc. Internat. Conf. on Automatic Face and Gesture Recognition (FGR'04)*, pp. 159–163, 2004. <https://doi.org/10.1109/AFGR.2004.1301524>
- [15] Costen, N., Brown, M., Akamatsu, S. "Sparse models for gender classification". In: *Proc. Internat. Conf. on Automatic Face and Gesture Recognition (FGR'04)*, pp. 201–206, 2004. <https://doi.org/10.1109/AFGR.2004.1301531>
- [16] Sun, N., Zheng, W., Sun, C., Zou, C., Zhao, L. "Gender classification based on boosting local binary pattern". In: *Proc. 3rd Internat. Symposium on Neural Networks (ISNN'06)*, Chengdu, China, vol. 2, pp. 194–201, 2006. https://doi.org/10.1007/11760023_29
- [17] Lian, H.-C., Lu, B.-L. "Multi-view gender classification using local binary patterns and support vector machines". In: *Proc. 3rd Internat. Sympos. on Neural Networks (ISNN'06)*, Chengdu, China, vol. 2, pp. 202–209, 2006. https://doi.org/10.1007/11760023_30
- [18] Baluja, S., Rowley, H.A. "Boosting sex identification performance". *Internat. J. Comput. Vision* vol. 71, no.1, pp.111–119, 2007. <https://doi.org/10.1007/s11263-006-8910-9>
- [19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks" in *Advances in neural information processing systems*, 2012, pp. 1097-1105, 2012. <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-network>
- [20] Y. Wen, Z., Li, Y. Qiao, "Latent Factor Guided Convolutional Neural Networks for Age-Invariant Face Recognition", in *IEEE Conference on Computer Vision and Pattern*

- Recognition, 2016. http://openaccess.thecvf.com/content_cvpr_2016/html/Wen_Latent_Factor_Guided_CVPR_2016_paper.html
- [21] F. Nian, L. Li, T. Li, “Robust gender classification on unconstrained face images”, in ACM International Conference on Internet Multimedia Computing and Service, 2015, pp. 77. <https://dl.acm.org/citation.cfm?id=2808570>
- [22] Z. Liu, P. Luo, X. Wang, and X. Tang, “Deep learning face attributes in the wild”, in IEEE International Conference on Computer Vision, 2015, pp. 3730-3738. http://openaccess.thecvf.com/content_iccv_2015/html/Liu_Deep_Learning_Face_ICCV_2015_paper.html
- [23] R. Olga, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, F. Li, “Imagenet large scale visual recognition challenge”, International Journal of Computer Vision, 2015, vol. 115, no. 3, pp. 211-252. <https://doi.org/10.1007/s11263-015-0816-y>
- [24] G. Levi, T. Hassner, “Age and Gender Classification Using Convolutional Neural Network”, in IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2015, pp. 34-42. https://www.cv-foundation.org/openaccess/content_cvpr_workshops_2015/W08/html/Levi_Age_and_Gender_2015_CVPR_paper.html
- [25] Eidinger, E., Enbar, R., Hassner, T. “Age and gender estimation of unfiltered faces”. In: Trans. on Inf. Forensics and Security. IEEE, pp. 2170–2179, 2014. <https://doi.org/10.1109/TIFS.2014.2359646>
- [26] Antipov, G., Berrani, S.A., Dugelay, J.L. “Minimalistic CNN-based ensemble model for gender prediction from face images”. In: Pattern Recognition Letters. S. 59-65, 2016. <https://doi.org/10.1016/j.patrec.2015.11.011>
- [27] Huang, G.B. “Labeled faces in the wild: A database for studying face recognition in unconstrained environments”. University of Massachusetts, Amherst. Pp. 7-49, 2007. <https://hal.inria.fr/inria-00321923/>
- [28] Mansanet, J., Albiol, A., Paredes, R. “Local deep neural networks for gender recognition”. In: Pattern Recognition Letters. pp. 80-86, 2016. <https://doi.org/10.1016/j.patrec.2015.11.015>
- [29] Pan, S.J. and Q. Yang. “A survey on transfer learning”. IEEE Transactions on Knowledge and Data Engineering, vol. 22, no.10, pp. 1345-1359, 2010. <https://doi.org/10.1109/TKDE.2009.191>
- [30] Hubel, D.H. and T.N. “Wiesel, Receptive fields, binocular interaction and functional architecture in the cat's visual cortex”. Journal of Physiology, vol. 160, no. 1, pp. 106-154, 1962. <https://doi.org/10.1113/jphysiol.1962.sp006837>
- [31] Ahmed, A., K. Yu, W. Xu, Y. Gong and E. Xing, “Training hierarchical feed-forward visual recognition models using transfer learning from pseudo-tasks”, in Computer Vision–ECCV, Springer. pp. 69-82, 2008. https://doi.org/10.1007/978-3-540-88690-7_6
- [32] Huang, J.T., J. Li, D. Yu, L. Deng and Y. Gong, “Cross-language knowledge transfer using multilingual deep neural network with shared hidden layers”. in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2013. <https://doi.org/10.1109/ICASSP.2013.6639081>