

Deep Residual Network and Transfer Learning-based Person Re-Identification

Fiola Angelica Deswita Allen, Tiffany Jacklin Polles, Evelyn Eleonora

Informatika, Siwalankerto No.121-131, Siwalankerto, Kec. Wonocolo, Surabaya, Jawa Timur

¹C14210021@john.petra.ac.id

²C14210056@john.petra.ac.id

³C14210119@john.petra.ac.id

Abstract— Kamera pengawas saat ini telah banyak digunakan di tempat umum untuk menjaga keamanan, untuk mengidentifikasi individu berdasarkan penampilan dan bentuk tubuh. Namun penerapan identifikasi individu masih menjadi suatu tantangan karena banyaknya variabel penghalang, seperti sudut pandang kamera, pencahayaan, halangan benda lain, dan lain sebagainya. Oleh karena itu, diperlukan teknologi canggih untuk identifikasi ulang individu (Re-ID). Penelitian ini mengusulkan model identifikasi individu menggunakan Residual Network (ResNet) dan Transfer Learning. ResNet, yang efektif dalam berbagai tugas Computer Vision, disesuaikan untuk identifikasi individu dengan mengeksplorasi beberapa variasi: ResNet-18, ResNet-34, ResNet-50, dan ResNeXt-50_32 \times 4D. Diterapkan metode Fine Tuning pada model yang telah dilakukan Transfer Learning untuk meningkatkan kemampuan model dalam mendeteksi ulang individu. Selain itu, dilakukan pengujian dengan beberapa dataset untuk membandingkan kemampuan adaptasi model pada situasi ketika mendeteksi ulang individu. Selain itu, diterapkan pula Gradient Scaling untuk mengoptimalkan kinerja model. Hasil dari eksperimen menunjukkan nilai akurasi dan kecepatan yang bagus dalam menjalankan tugas deteksi individu.

Keywords— Person Re-Identification, Transfer Learning, Residual Network, Fine Tuning, Gradient Scaling

I. INTRODUCTION

Kamera pengawas sudah marak digunakan untuk menjaga keamanan dan ketertiban di berbagai tempat umum, seperti bandara, sekolah, kampus, dan tempat umum lainnya. Salah satu kegunaan dari kamera pengawas adalah untuk mengidentifikasi seseorang berdasarkan penampilan dan bentuk tubuh. Agar kamera dapat mengenali seseorang, maka diperlukan suatu teknologi yang dapat melakukan deteksi. Penggunaan teknologi untuk Re-ID ini dapat mengenali seseorang di tempat dan waktu yang berbeda, serta dari sudut kamera yang berbeda [1]. Namun, terdapat beberapa tantangan utama dalam pengidentifikasian seseorang melalui kamera [2], yaitu adanya variasi dalam sudut pandang kamera, pencahayaan, pose manusia, halangan benda lain, resolusi yang rendah, serta latar belakang yang berbeda-beda.

Identifikasi seseorang merupakan bentuk permasalahan *open-world*, di mana terdapat kemungkinan orang yang ingin diidentifikasi tidak ada di dalam dataset. Hal ini mengharuskan untuk memperbanyak jumlah *class* dalam dataset untuk melakukan deteksi [3], [4]. Dalam hal dataset,

pengumpulan data untuk tugas identifikasi juga merupakan hal yang sulit untuk dilakukan. Pengumpulan identitas untuk data membutuhkan waktu dan tenaga yang cukup banyak untuk membuat dataset dalam jumlah besar. Hal ini menyebabkan dataset memiliki kekurangan dalam ukuran dataset yang berlabel [5]. Oleh karena itu, pada penelitian ini akan dilakukan *transfer learning* dimana model akan di *training* pada dataset yang lebih besar, kemudian model akan disesuaikan untuk menyesuaikan dengan dataset yang digunakan. Dataset yang digunakan pada paper ini adalah ImageNet dimana dataset ini sudah terbukti sukses ketika digunakan untuk transfer learning.

II. DATASETS

Penelitian ini akan menguji model menggunakan 3 tipe dataset.

A. CUHK03-NP (*Chinese University of Hong Kong Re-identification*)

Dataset CUHK03-NP [6] ini memiliki jumlah gambar sebanyak 14,097 dengan 1,467 identitas yang diambil dengan menggunakan dua buah kamera. Tiap gambarnya memiliki resolusi 640x480.

B. Market-1501

Dataset Market-1501 [7] dikumpulkan di depan supermarket di Universitas Tsinghua. Gambar yang diambil berasal dari satu kamera resolusi rendah dan enam kamera resolusi tinggi. Dataset ini memiliki jumlah gambar sebanyak 32,668 dengan 1,501 identitas.

C. Combined (*Gabungan Dataset CUHK03 dan Market-1501*)

Dataset CUHK03-NP dan Market-1501 kemudian digabung untuk membentuk dataset baru. Penggabungan kedua dataset ini dilakukan agar memiliki dataset dalam jumlah yang besar.

D. Dataset Baru

Dataset ini berasal dari film, drama dan video yang dikumpulkan oleh tiap anggota. Gambar yang diambil memiliki resolusi yang berbeda-beda. Dataset ini memiliki jumlah gambar sebanyak 390 dengan 60 identitas. Dataset yang dikumpulkan mandiri ini hanya digunakan untuk melakukan pengujian pada model yang telah di-training

dengan dataset CUHK03-NP, Market-1501, dan Combined. Setiap identitas akan diambil satu buah gambar yang akan digunakan sebagai data query, sedangkan sisanya sebagai data gallery. Pembagian data query dan gallery ini digunakan untuk perhitungan nilai evaluasi.

Pada dataset CUHK03-NP, Market-1501, dan Combined dilakukan dua buah percobaan dimana percobaan pertama dilakukan train-test split dengan perbandingan 50:50 dan percobaan kedua dengan perbandingan 80:20. Kemudian, diambil satu buah gambar pada setiap identitas dalam data training untuk dijadikan sebagai data validasi. Setiap gambar dalam dataset kemudian diubah resolusinya menjadi 256x128 dan dilakukan normalisasi.

III. PROPOSED METHOD

Penelitian ini mengusulkan model pengenalan individu yang memanfaatkan *residual network* (ResNet) dan teknik *transfer learning*. ResNet telah terbukti efektif dalam berbagai tugas *computer vision*, dan model ini dapat mempelajari fitur-fitur dari dataset pelatihan yang besar dan kemudian disesuaikan pada dataset pengenalan individu. Penelitian ini mengeksplorasi beberapa variasi ResNet, seperti ResNet-18, ResNet-34, ResNet-50, dan ResNeXt-50_32 \times 4D, untuk menentukan model yang paling cocok untuk tugas pengenalan individu. Teknik *fine-tuning* juga digunakan untuk menyesuaikan model *pre-trained* pada dataset besar ke dataset pengenalan individu. Dengan memanfaatkan *transfer learning* dan *fine-tuning*, diharapkan model yang diusulkan dapat mengatasi keterbatasan dataset berlabel dalam pengenalan individu dan meningkatkan kinerja identifikasi individu melalui jaringan kamera.

A. ResNet-18

ResNet-18 adalah model dengan 18 lapisan yang terdiri dari 4 blok residual, masing-masing dengan serangkaian lapisan konvolusi diikuti oleh *batch normalization* dan aktivasi ReLU. Keberadaan *shortcut connection* dalam setiap blok residual memungkinkan aliran gradien yang lancar, mengatasi masalah gradien yang menghilang, dan memungkinkan pelatihan model yang lebih dalam. Meskipun ResNet-18 cenderung memberikan hasil yang akurat dalam masalah Re-ID karena kemampuannya mempelajari fitur-fitur yang relevan, tingkat *error* yang tinggi masih mungkin terjadi, tergantung pada faktor-faktor seperti kompleksitas masalah dan kualitas data, yang dapat diatasi dengan teknik-teknik penyesuaian dan pengoptimalan yang tepat [5].

B. ResNet-34

ResNet-34 adalah model dengan 34 lapisan yang mencakup lapisan konvolusi, lapisan *max-pooling*, 4 blok residual, lapisan *average pooling*, *fully connected*, dan *SoftMax*. Setiap blok residual menggunakan filter 3x3 yang diatur dalam kelompok 3, 4, 6, dan 3. Dibandingkan dengan ResNet-18, ResNet-34 cenderung memiliki training *error* yang lebih rendah dan kemampuan generalisasi yang lebih baik. Namun, dalam beberapa penelitian, performa ResNet-18 dan ResNet-34 sering kali sangat mirip, menunjukkan bahwa

meskipun ResNet-34 memiliki arsitektur yang lebih dalam, keuntungannya dalam hal performa tidak selalu signifikan [5].

C. ResNet-50

ResNet-50 adalah model dengan 50 lapisan yang terdiri dari 4 blok residual. Berbeda dengan ResNet-34, model ini menggantikan dua blok filter pada ResNet-34 dengan blok 3-lapisan, yang terdiri dari filter 1x1, filter 3x3, dan filter 1x1 lagi. Penggunaan filter 1x1 bertujuan untuk mengurangi dimensi data sebelumnya dan kemudian mengembalikannya, sehingga lapisan 3x3 akan memiliki input dan output dengan dimensi yang lebih kecil, memperkaya representasi fitur tanpa menambah jumlah parameter secara signifikan [5].

D. ResNeXt-50_32 \times 4D

ResNeXt-50_32 \times 4D adalah jaringan yang menggabungkan elemen-elemen dari VGG dan modifikasi dari ResNet-50. Dalam model ini, terdapat tumpukan blok residual yang mengikuti dua aturan penting: pertama, kedua blok harus memiliki *hyperparameter* yang sama jika menghasilkan *spatial map* dengan ukuran yang sama, dan kedua, lebar blok akan dikalikan dengan faktor 2 ketika ukuran *spatial map* berkurang setengahnya, menciptakan dimensi yang baru [5].

Salah satu inovasi utama dari ResNeXt adalah pengenalan dimensi baru yang disebut *cardinality*. *Cardinality* mengacu pada jumlah transformasi yang digabungkan dalam jaringan. Semakin besar nilai *cardinality*, semakin banyak transformasi yang digabungkan [5]. Penelitian sebelumnya menunjukkan bahwa penggunaan *cardinality* lebih efektif daripada menambahkan kedalaman (*depth*) atau lebar (*width*) pada jaringan, membantu dalam memperkaya representasi fitur tanpa menambahkan kompleksitas yang berlebihan [8].

E. Transfer Learning

Transfer learning digunakan karena minimnya dataset yang memiliki label untuk melatih *person re-identification* (re-id), serta kesulitan dalam mengumpulkan dataset yang diperlukan untuk tugas ini. Oleh karena itu, pendekatan *transfer learning* dilakukan dengan menggunakan dataset yang lebih besar, yaitu ImageNet, sebagai tahap *pre-training* model. Teknik *transfer learning* ini sangat bermanfaat pada dataset yang tidak memiliki label dan memiliki ukuran yang besar.

F. Fine Tuning

Fine-tuning (FT) adalah desain populer dalam transfer learning di mana jaringan pertama kali dipre-train pada dataset yang lebih besar. Kemudian, beberapa lapisan diekstraksi dari model yang dipre-train sesuai dengan jaringan target, dan baik semua atau sebagian dari lapisan tersebut disesuaikan ulang untuk mendapatkan jaringan yang diinginkan. Meningkatkan kedalaman jaringan tidak efektif, meskipun penyesuaian transformasi lapisan dapat meningkatkan kinerja [5].

G. Gradient Scaling

Gradient scaling (GS) adalah sebuah teknik yang penting dalam optimasi model *deep learning*, telah diterapkan dalam berbagai tugas, mulai dari meningkatkan kinerja jaringan saraf

dalam pengenalan gambar hingga memastikan efisiensi model dalam sistem real-time. Implementasi gradient scaling merupakan langkah kritis untuk mencapai efisiensi dan stabilitas yang lebih tinggi selama proses pelatihan [8].

IV. EXPERIMENT

Eksperimen yang dilakukan pada penelitian ini adalah penerapan dataset CUHK03-NP, Market-1501, dan Combined pada beberapa variasi model ResNet. Dataset ini juga akan dilakukan eksperimen pada pembagian data training dan data testing dengan perbandingan 50:50 dan 80:20. Selain itu, dilakukan eksperimen dengan menerapkan GS pada setiap model. Model juga akan dilakukan fine-tuning, dengan mengubah nilai parameter, untuk menguji peningkatan nilai akurasi hasil deteksi. Setiap model ResNet yang diuji akan diterapkan metode TF dengan menggunakan nilai *weight* dari hasil training yang telah dilakukan pada dataset ImageNet. Hasil dari eksperimen ini akan melihat perbandingan hasil deteksi Re-ID pada setiap model ResNet.

Evaluasi model yang digunakan adalah perhitungan nilai mAP dan CMC yang diambil top-1, top-5, dan top-10. Data *testing* terlebih dahulu akan dijadikan menjadi beberapa *batch* dan dilakukan ekstraksi fitur dan label pada semua gambar dalam data *gallery* dan *query*, serta diterapkan L2 Normalization. Fitur-fitur yang telah diekstrak dan dinormalisasi kemudian akan dilakukan perhitungan jarak

menggunakan *Euclidean Distance*. Perhitungan jarak dihitung antara satu data *query* pada setiap data *gallery*. Dari nilai jarak ini kemudian dilakukan perhitungan nilai mAP dan CMC untuk evaluasi model.

A. Without Fine-Tuning

Pada eksperimen ini dilakukan *training* pada semua variasi model ResNet dengan mengambil *weight* dari ImageNet. Eksperimen ini menggunakan *optimizer* Adam dengan *learning rate* sebesar 0.001. Dari hasil *training* yang ditunjukkan pada Tabel 1, dapat dilihat bahwa model ResNet-18 dengan GS dan perbandingan train-test split sebesar 50:50, yang diuji pada dataset Market-1501 mencapai nilai akurasi tertinggi, yaitu 1.000 dan loss terkecil yaitu 0.0038. Nilai akurasi validasi yang diperoleh oleh model ResNet-18 dengan GS juga mencapai nilai tertinggi dibandingkan model lainnya, yaitu 0.8895.

Hasil pengujian ditunjukkan pada Table 2, di mana pengujian dilakukan menggunakan data *testing* dari train-test split dengan perbandingan 50:50 dan 80:20. Jika dilihat pada Table 2, nilai CMC Top-1 dan CMC Top-5 tertinggi dimiliki oleh Model ResNet-18 dengan nilai sebesar 0.9733 dan 0.9903. CMC Top-10 terbaik dimiliki oleh model ResNet-18 dan ResNet-34, yaitu 0.9941. Sedangkan, nilai mAP tertinggi didapat oleh Model ResNeXt-50_32x4D, yaitu 0.7156.

TABLE I
TRAINING WITHOUT FINE-TUNING

| Model | Dataset | Training: Testing | Without Fine-Tuning | | | | | | | |
|----------------------|-------------|----------------------|---------------------|----------|--------|-----------------|-----------------|---------------|---------------|---------------|
| | | | Without GradScaler | | | | With GradScaler | | | |
| | | | Best Val Acc | Accuracy | Loss | Time | Best Val Acc | Accuracy | Loss | Time |
| ResNet-18 | CUHK03-NP | 50:50 | 0.5254 | 0.9992 | 0.1666 | 15m 19s | 0.3937 | 0.9971 | 0.2591 | 14m 51s |
| | | 80:20 | 0.4893 | 0.9912 | 0.3926 | 20m 59s | 0.5286 | 0.9834 | 0.4425 | 19m 28s |
| | MARKET-1501 | 50:50 | 0.8575 | 1.0000 | 0.0057 | 21m 37s | 0.8895 | 1.0000 | 0.0038 | 17m 41s |
| | | 80:20 | 0.8285 | 1.0000 | 0.0039 | 39m 56s | 0.8193 | 1.0000 | 0.0032 | 29m 23s |
| | COMBINED | 50:50 | 0.5256 | 1.0000 | 0.0112 | 28m 59s | 0.8095 | 1.0000 | 0.0085 | 22m 48s |
| | | 80:20 | 0.4701 | 1.0000 | 0.0070 | 60m 30s | 0.7860 | 1.0000 | 0.0075 | 48m 41s |
| ResNet-34 | CUHK03-NP | 50:50 | 0.3781 | 0.9912 | 0.3395 | 19m 41s | 0.4876 | 0.9979 | 0.2186 | 15m 55s |
| | | 80:20 | 0.2907 | 0.9799 | 0.4300 | 27m 48s | 0.4399 | 0.9865 | 0.3743 | 21m 24s |
| | MARKET-1501 | 50:50 | 0.8056 | 1.0000 | 0.0096 | 32m 16s | 0.8176 | 1.0000 | 0.0158 | 19m 27s |
| | | 80:20 | 0.8018 | 1.0000 | 0.0102 | 59m 12s | 0.8135 | 1.0000 | 0.0074 | 35m 55s |
| | COMBINED | 50:50 | 0.4771 | 1.0000 | 0.0170 | 41m 52s | 0.7090 | 1.0000 | 0.0144 | 26m 9s |
| | | 80:20 | 0.7152 | 1.0000 | 0.0108 | 84m 52s | 0.7056 | 1.0000 | 0.0103 | 53m 46s |
| ResNet-50 | CUHK03-NP | 50:50 | 0.3168 | 0.8125 | 1.1420 | 32m 36s | 0.4915 | 0.9344 | 0.5648 | 19m 23s |
| | | 80:20 | 0.2864 | 0.7343 | 1.4151 | 45m 24s | 0.2472 | 0.6737 | 1.6417 | 27m 16s |
| | MARKET-1501 | 50:50 | 0.5193 | 0.9977 | 0.0479 | 30m 2s | 0.6352 | 0.9990 | 0.0308 | 29m 53s |
| | | 80:20 | 0.5279 | 0.9982 | 0.0410 | 95m 25s | 0.5262 | 0.9989 | 0.0403 | 47m 16s |
| | COMBINED | 50:50 | 0.4877 | 0.9995 | 0.0316 | 39m 6s | 0.5256 | 0.9950 | 0.0832 | 38m 49s |
| | | 80:20 | 0.4899 | 0.9995 | 0.0290 | 72m 59s | 0.4701 | 0.9993 | 0.0312 | 72m 19s |
| ResNeXt-50 _32x4D | CUHK03-NP | 50:50 | 0.3064 | 0.9133 | 1.0120 | 42m 19s | 0.4342 | 0.9610 | 0.3259 | 21m 38s |
| | | 80:20 | 0.1654 | 0.8487 | 1.2283 | 59m 26s | 0.2643 | 0.9806 | 0.3418 | 30m 5s |
| | MARKET-1501 | 50:50 | 0.7071 | 0.9995 | 0.0398 | 38m 34s | 0.6218 | 0.9997 | 0.0394 | 36m 27s |
| | | 80:20 | 0.7294 | 0.9998 | 0.0172 | 119m 59s | 0.6653 | 0.9996 | 0.0333 | 58m 8s |
| | COMBINED | 50:50 | 0.7222 | 0.9995 | 0.0316 | 47m 34s | 0.4771 | 0.9966 | 0.0853 | 47m 39s |
| | | 80:20 | 0.5661 | 0.9995 | 0.0290 | 88m 9s | 0.7152 | 0.9999 | 0.0184 | 89m 15s |

TABLE II
TESTING WITHOUT FINE-TUNING

| Model | Dataset | Training: Testing | Without Fine-Tuning | | | | | | | |
|------------------|-------------|----------------------|---------------------|--------|--------|-----------------|--------|--------|--------|--------|
| | | | Without GradScaler | | | With GradScaler | | | | |
| | | | mAP | CMC | | mAP | CMC | | Top-1 | |
| ResNet-18 | CUHK03-NP | 50:50 | 0.3626 | 0.7407 | 0.8886 | 0.9364 | 0.3639 | 0.7221 | 0.8836 | 0.9193 |
| | | 80:20 | 0.4698 | 0.8503 | 0.9541 | 0.9694 | 0.4663 | 0.8248 | 0.9473 | 0.9694 |
| | MARKET-1501 | 50:50 | 0.0043 | 0.0003 | 0.0048 | 0.0086 | 0.0045 | 0.0003 | 0.0050 | 0.0101 |
| | | 80:20 | 0.6830 | 0.9733 | 0.9903 | 0.9941 | 0.6888 | 0.9718 | 0.9903 | 0.9941 |
| | COMBINED | 50:50 | 0.0028 | 0.0006 | 0.0040 | 0.0076 | 0.0026 | 0.0004 | 0.0034 | 0.0073 |
| | | 80:20 | 0.5816 | 0.8164 | 0.9240 | 0.9493 | 0.6598 | 0.8548 | 0.9499 | 0.9719 |
| ResNet-34 | CUHK03-NP | 50:50 | 0.3482 | 0.7164 | 0.8750 | 0.9157 | 0.3607 | 0.7257 | 0.8950 | 0.9343 |
| | | 80:20 | 0.4561 | 0.8197 | 0.9354 | 0.9660 | 0.4428 | 0.7908 | 0.9252 | 0.9626 |
| | MARKET-1501 | 50:50 | 0.0046 | 0.0003 | 0.0059 | 0.0107 | 0.0050 | 0.0003 | 0.0062 | 0.0122 |
| | | 80:20 | 0.3848 | 0.9598 | 0.9821 | 0.9896 | 0.7016 | 0.9673 | 0.9889 | 0.9941 |
| | COMBINED | 50:50 | 0.0025 | 0.0004 | 0.0027 | 0.0078 | 0.0028 | 0.0006 | 0.0040 | 0.0076 |
| | | 80:20 | 0.5563 | 0.7916 | 0.9157 | 0.9504 | 0.5816 | 0.8164 | 0.9240 | 0.9493 |
| ResNet-50 | CUHK03-NP | 50:50 | 0.2552 | 0.5579 | 0.7621 | 0.8343 | 0.3072 | 0.6664 | 0.8450 | 0.9057 |
| | | 80:20 | 0.3596 | 0.6667 | 0.8605 | 0.9133 | 0.3309 | 0.6207 | 0.8367 | 0.9065 |
| | MARKET-1501 | 50:50 | 0.0047 | 0.0003 | 0.0059 | 0.0137 | 0.0052 | 0.0003 | 0.0089 | 0.0187 |
| | | 80:20 | 0.5981 | 0.9509 | 0.9799 | 0.9896 | 0.5983 | 0.9449 | 0.9792 | 0.9859 |
| | COMBINED | 50:50 | 0.0029 | 0.0004 | 0.0048 | 0.0113 | 0.0027 | 0.0004 | 0.0050 | 0.0096 |
| | | 80:20 | 0.4373 | 0.7081 | 0.8559 | 0.9082 | 0.4164 | 0.6827 | 0.8440 | 0.8937 |
| ResNeXt-50_32x4D | CUHK03-NP | 50:50 | 0.2076 | 0.4193 | 0.6643 | 0.7693 | 0.2652 | 0.5550 | 0.7793 | 0.8529 |
| | | 80:20 | 0.2561 | 0.4592 | 0.7466 | 0.8299 | 0.4210 | 0.7007 | 0.8929 | 0.9422 |
| | MARKET-1501 | 50:50 | 0.0048 | 0.0009 | 0.0068 | 0.0125 | 0.0051 | 0.0009 | 0.0065 | 0.0119 |
| | | 80:20 | 0.7156 | 0.9665 | 0.9874 | 0.9918 | 0.6544 | 0.9360 | 0.9784 | 0.9859 |
| | COMBINED | 50:50 | 0.0022 | 0.0004 | 0.0025 | 0.0061 | 0.0021 | 0.0000 | 0.0019 | 0.0042 |
| | | 80:20 | 0.5090 | 0.7538 | 0.8953 | 0.9285 | 0.5643 | 0.8051 | 0.9248 | 0.9502 |

Dari hasil pengujian ini, seluruh nilai terbaik didapat pada model yang ditraining menggunakan perbandingan train-test split 80:20. Seluruh model yang ditraining menggunakan train-test split dengan perbandingan 50:50 memiliki akurasi yang jauh lebih rendah dibandingkan model yang ditraining menggunakan train-test split dengan perbandingan 80:20.

B. With Fine-Tuning

Pada eksperimen ini dilakukan training dengan menggunakan model serta dataset yang sama dengan eksperimen tanpa FT. Tetapi, dikarenakan mayoritas nilai akurasi pada train-test split dengan perbandingan 50:50 menghasilkan nilai yang cukup rendah, maka pada percobaan dengan FT hanya dilakukan pada dataset dengan perbandingan 80:20. *Fine-Tuning* yang dilakukan pada eksperimen ini adalah perubahan nilai parameter pada *optimizer* yang digunakan. Parameter yang digunakan adalah *learning rate* sebesar 2e-4 dan *weight decay* sebesar 0.015. Dari hasil *training* yang ditunjukkan pada Table 3, dapat dilihat bahwa model ResNeXt-50_32x4D tanpa GS yang diuji pada dataset Combined memperoleh nilai akurasi *training* tertinggi, yaitu 1.000 dengan nilai loss sebesar 0.0028. Tetapi, model ResNeXt-50_32x4D dengan GS yang diuji pada dataset CUHK03-NP mencapai nilai akurasi validasi terbaik, yaitu sebesar 0.9661.

Dari hasil pengujian yang dapat dilihat pada Table 4, model ResNet-50 dengan GS yang diuji pada dataset Market-1501

mencapai nilai CMC tertinggi, yaitu Top-1 sebesar 0.9911, Top-5 sebesar 0.9963. Sedangkan CMC Top-10 tertinggi dipegang oleh model ResNet-34 dengan GS yang diuji pada dataset Market-1501, yaitu sebesar 0.9985. Namun nilai mAP tertinggi dimiliki oleh model ResNeXt-50_32x4D tanpa GS yang diuji pada dataset Market-1501, yaitu sebesar 0.8995.

Pengujian selanjutnya yang ditunjukkan pada Table 5 dilakukan dengan menggunakan dataset baru. Pada percobaan ini, nilai mAP tertinggi tetap dimiliki oleh Model ResNeXt-50_32x4D, yaitu 0.7145. Model ResNeXt-50_32x4D juga memiliki CMC Top-1 tertinggi dibanding model lainnya, yaitu 0.9167. Nilai CMC Top-5 dan CMC Top-10 tertinggi dimiliki oleh Model ResNet-34 dan ResNet-50 dengan nilai sebesar 0.9833.

C. Gradient Scaling

Penggunaan GS pada model menunjukkan dampak yang cukup signifikan jika dibandingkan tanpa menggunakan GS. Secara keseluruhan, penggunaan GS pada model memberikan dampak pada kecepatan deteksi yang dilakukan menjadi lebih cepat. Dapat dilihat pada Table 1, sebelum dilakukan fine-tuning, model ResNeXt-50_32x4D yang menggunakan dataset Market-1501 dengan train-test split 80:20 memiliki durasi paling panjang yaitu 119m 59s ketika tidak menggunakan GS, dan menjadi 58m 8s ketika menggunakan GS. Selisih antara kedua durasi ini mencapai 61m 51s.

TABLE III
TRAINING WITH FINE-TUNING

| Model | Dataset | Training: Testing | With Fine-Tuning | | | | | | | |
|----------------------|-------------|----------------------|--------------------|----------|--------|-----------------|-----------------|---------------|---------------|----------------|
| | | | Without GradScaler | | | | With GradScaler | | | |
| | | | Best Val Acc | Accuracy | Loss | Time | Best Val Acc | Accuracy | Loss | Time |
| ResNet-18 | CUHK03-NP | 80:20 | 0.8827 | 1.0000 | 0.0227 | 15m 53s | 0.8644 | 1.0000 | 0.0223 | 14m 53s |
| | MARKET-1501 | 80:20 | 0.8526 | 1.0000 | 0.0115 | 33m 7s | 0.8518 | 1.0000 | 0.0113 | 24m 53s |
| | COMBINED | 80:20 | 0.8854 | 1.0000 | 0.0103 | 53m 3s | 0.8766 | 1.0000 | 0.0110 | 43m 51s |
| ResNet-34 | CUHK03-NP | 80:20 | 0.9400 | 1.0000 | 0.0118 | 20m 1s | 0.9309 | 1.0000 | 0.0127 | 16m 19s |
| | MARKET-1501 | 80:20 | 0.9117 | 1.0000 | 0.0056 | 50m 42s | 0.9092 | 1.0000 | 0.0056 | 29m 1s |
| | COMBINED | 80:20 | 0.9414 | 1.0000 | 0.0054 | 77m 36s | 0.9132 | 1.0000 | 0.0065 | 48m 51s |
| ResNet-50 | CUHK03-NP | 80:20 | 0.9544 | 1.0000 | 0.8500 | 34m 39s | 0.2295 | 1.0000 | 0.0625 | 27m 10s |
| | MARKET-1501 | 80:20 | 0.9092 | 1.0000 | 0.0046 | 94m 38s | 0.9209 | 1.0000 | 0.0045 | 46m 42s |
| | COMBINED | 80:20 | 0.8715 | 1.0000 | 0.0069 | 71m 30s | 0.8854 | 1.0000 | 0.0055 | 72m 37s |
| ResNeXt-50- 32x4D | CUHK03-NP | 80:20 | 0.9570 | 1.0000 | 0.0067 | 43m 27s | 0.9661 | 1.0000 | 0.0064 | 21m 42s |
| | MARKET-1501 | 80:20 | 0.9217 | 1.0000 | 0.0030 | 125m 13s | 0.9276 | 1.0000 | 0.0028 | 57m 57s |
| | COMBINED | 80:20 | 0.9377 | 1.0000 | 0.0025 | 91m 31s | 0.9414 | 1.0000 | 0.0028 | 87m 53s |

TABLE IV
TESTING WITH FINE-TUNING

| Model | Dataset | Training: Testing | With Fine-Tuning | | | | | | | |
|----------------------|-------------|----------------------|--------------------|--------|--------|--------|-----------------|---------------|---------------|---------------|
| | | | Without GradScaler | | | | With GradScaler | | | |
| | | | mAP | CMC | | | mAP | CMC | | |
| ResNet-18 | CUHK03-NP | 80:20 | 0.4208 | 0.8521 | 0.9450 | 0.9657 | 0.4128 | 0.8550 | 0.9393 | 0.9593 |
| | MARKET-1501 | 80:20 | 0.7834 | 0.9821 | 0.9918 | 0.9955 | 0.7892 | 0.9807 | 0.9933 | 0.9940 |
| | COMBINED | 80:20 | 0.5787 | 0.8169 | 0.9285 | 0.9522 | 0.5859 | 0.8323 | 0.9336 | 0.9552 |
| ResNet-34 | CUHK03-NP | 80:20 | 0.5664 | 0.9193 | 0.9750 | 0.9850 | 0.5676 | 0.9307 | 0.9750 | 0.9864 |
| | MARKET-1501 | 80:20 | 0.8689 | 0.9903 | 0.9948 | 0.9978 | 0.8651 | 0.9881 | 0.9963 | 0.9985 |
| | COMBINED | 80:20 | 0.6814 | 0.8776 | 0.9568 | 0.9697 | 0.6667 | 0.8596 | 0.9455 | 0.9645 |
| ResNet-50 | CUHK03-NP | 80:20 | 0.6382 | 0.9229 | 0.9779 | 0.9879 | 0.6678 | 0.9422 | 0.9881 | 0.9949 |
| | MARKET-1501 | 80:20 | 0.8734 | 0.9859 | 0.9948 | 0.9955 | 0.8837 | 0.9911 | 0.9963 | 0.9970 |
| | COMBINED | 80:20 | 0.6542 | 0.8544 | 0.9398 | 0.9604 | 0.6675 | 0.8585 | 0.9434 | 0.9655 |
| ResNeXt-50- 32x4D | CUHK03-NP | 80:20 | 0.6133 | 0.9393 | 0.9779 | 0.9893 | 0.6137 | 0.9393 | 0.9800 | 0.9907 |
| | MARKET-1501 | 80:20 | 0.8995 | 0.9888 | 0.9970 | 0.9978 | 0.8943 | 0.9859 | 0.9955 | 0.9970 |
| | COMBINED | 80:20 | 0.8592 | 0.9560 | 0.9824 | 0.9886 | 0.8636 | 0.9488 | 0.9840 | 0.9896 |

TABLE V
TESTING WITH OUR DATASET AND FINE-TUNED MODEL

| Model | Dataset | With Fine-Tuning | | | | | | | | |
|----------------------|-------------|--------------------|---------------|---------------|---------------|-----------------|--------|--------|--------|--------|
| | | Without GradScaler | | | | With GradScaler | | | | |
| | | mAP | CMC | | | mAP | CMC | | | Top-10 |
| ResNet-18 | CUHK03-NP | 0.4520 | 0.7167 | 0.8833 | 0.9667 | 0.4672 | 0.6667 | 0.8500 | 0.9000 | |
| | MARKET-1501 | 0.5595 | 0.7500 | 0.9000 | 0.9667 | 0.5567 | 0.8000 | 0.9833 | 0.9833 | |
| | COMBINED | 0.5527 | 0.8000 | 0.9167 | 0.9667 | 0.5650 | 0.7500 | 0.9167 | 0.9833 | |
| ResNet-34 | CUHK03-NP | 0.5674 | 0.7333 | 0.9333 | 0.9667 | 0.5273 | 0.7333 | 0.9667 | 0.9833 | |
| | MARKET-1501 | 0.6371 | 0.8500 | 0.9667 | 0.9833 | 0.6355 | 0.7833 | 0.9167 | 0.9833 | |
| | COMBINED | 0.6917 | 0.9000 | 0.9833 | 0.9833 | 0.6954 | 0.9000 | 0.9667 | 0.9667 | |
| ResNet-50 | CUHK03-NP | 0.4453 | 0.6000 | 0.8500 | 0.9500 | 0.3346 | 0.5167 | 0.8167 | 0.8167 | |
| | MARKET-1501 | 0.6423 | 0.8500 | 0.9500 | 0.9667 | 0.6701 | 0.8667 | 0.9667 | 0.9833 | |
| | COMBINED | 0.6727 | 0.8833 | 0.9833 | 0.9833 | 0.6394 | 0.8167 | 0.9333 | 0.9833 | |
| ResNeXt-50- 32x4D | CUHK03-NP | 0.5930 | 0.7833 | 0.9333 | 0.9667 | 0.5493 | 0.8000 | 0.9333 | 0.9667 | |
| | MARKET-1501 | 0.7145 | 0.9167 | 0.9667 | 0.9667 | 0.6934 | 0.8833 | 0.9500 | 0.9500 | |
| | COMBINED | 0.6964 | 0.9000 | 0.9667 | 0.9667 | 0.6977 | 0.8667 | 1.0000 | 1.0000 | |

Setelah dilakukan fine-tuning, dapat dilihat pada Table 3, model ResNeXt-50_32x4D yang menggunakan dataset Market-1501 dengan train-test split 80:20 juga memiliki durasi paling panjang yaitu 125m 13s. Model ini mengalami

penurunan durasi yang sangat signifikan menjadi 57m 57s. Selisih antara kedua durasi ini mencapai 67m 16s. Hal ini menunjukkan penerapan GS pada model dapat meningkatkan kecepatan deteksi model.

Secara keseluruhan model ResNeXt-50_32x4D dengan TF, FT, dan GS memiliki performa terbaik. Eksperimen yang dilakukan terhadap berbagai variasi ResNet, yaitu ResNet-18, ResNet-34, ResNet-50, dan ResNeXt-50_32x4D, membuktikan bahwa model dengan layer yang lebih dalam tidak selalu memiliki performa lebih baik daripada model dengan layer yang lebih sedikit. Selain itu, penerapan FT dan penggunaan dataset berukuran besar mampu meningkatkan performa model dalam melakukan deteksi ulang individu. Model dapat mempelajari fitur-fitur dengan lebih optimal dan bervariasi sehingga mampu mengurangi terjadinya *overfitting*.

V. CONCLUSION

Berdasarkan eksperimen yang telah kami lakukan pada berbagai variasi ResNet dan dataset, model ResNet-18 dengan *Gradient Scaling* (GS) menunjukkan akurasi tertinggi pada dataset Market-1501 dengan pembagian 50:50, sementara model dengan pembagian 80:20 menunjukkan performa lebih baik secara keseluruhan dalam hal akurasi dan mAP. Penggunaan GS pada model mempersingkat durasi waktu deteksi sekaligus meningkatkan akurasi dan efisiensi model. Peningkatan ini menunjukkan bahwa GS merupakan teknik yang menguntungkan dalam pelatihan model. Penerapan fine-tuning pada model dapat memberikan peningkatan yang signifikan pada performa model. Dengan penerapan FT, model lebih mampu untuk mempelajari fitur-fitur sehingga terjadi peningkatan nilai akurasi pelatihan dengan menurunkan nilai loss. Model ResNeXt-50_32x4D menonjol dengan mAP tertinggi dengan kemampuan yang mengesankan dalam identifikasi ulang individu, terbukti dari nilai CMC Top-1 sebesar 0,9167 dan nilai yang memuaskan untuk Top-5 dan Top-10. Model ResNet-18 dan ResNeXt-50_32x4D, dengan *Gradient Scaling* dan pembagian *train-test* yang tepat,

menunjukkan kemampuan terbaik dalam menghasilkan nilai akurasi dan mAP, serta efisiensi komputasi untuk tugas identifikasi ulang individu.

REFERENCES

- [1] M. Ye, J. Shen, G. Lin, T. Xiang, L. Shao, and S. C. H. Hoi, "Deep Learning for Person Re-Identification: A Survey and outlook," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 6, pp. 2872–2893, Jun. 2022, doi: 10.1109/tpami.2021.3054775.
- [2] R. Zhao, W. Ouyang and X. Wang, "Unsupervised Salience Learning for Person Re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Portland, OR, USA, 2013, pp. 3586-3593, doi: 10.1109/CVPR.2013.460.
- [3] A. Bedagkar-Gala and S. K. Shah, "A survey of approaches and trends in person re-identification," *Image and Vision Computing*, vol. 32, no. 4, pp. 270–286, Apr. 2014, doi: 10.1016/j.imavis.2014.02.001.
- [4] L. Zheng, Y. Yang, and A. G. Hauptmann, "Person re-identification: past, present and future," arXiv (Cornell University), Jan. 2016, doi: 10.48550/arxiv.1610.02984.
- [5] A. Gupta, P. Pawade, and R. Balakrishnan, "Deep residual network and transfer learning-based person Re-Identification," *Intelligent Systems With Applications*, vol. 16, p. 200137, Nov. 2022, doi: 10.1016/j.iswa.2022.200137.
- [6] Z. Zhong, L. Zheng, D. Cao, and S. Li, "Re-ranking person re-identification with k-reciprocal encoding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017. Available: https://github.com/zhunzhong07/person-re-ranking/tree/master/CUHK_03-NP. [Accessed: Jun. 12, 2024].
- [7] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015. Available: https://zheng-lab.cecs.anu.edu.au/Project/project_reid.html. [Accessed: Jun. 12, 2024].
- [8] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-Excitation networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 8, pp. 2011–2023, Aug. 2020, doi: 10.1109/tpami.2019.2913372.
- [9] "Gradient Scaling" | DeepGram. Available: <https://deepgram.com/ai-glossary/gradient-scaling>. [Accessed: Jun. 26, 2024].

LAMPIRAN

CONTOH VISUALISASI DATASET CUHK03



CONTOH VISUALISASI DATASET MARKET-1501



CONTOH VISUALISASI DATASET BARU



CONTOH VISUALISASI TESTING

