Analisis Dampak Modifikasi Arsitektur pada Self-Supervised Learning SimCLR

Ghulam Mushthofa
442023611060
Universitas Darussalam Gontor
Surabaya, Indonesia
ghulammushthofa@gmail.com

***Abstract*—Penelitian ini bertujuan untuk menganalisis dampak modifikasi arsitektur pada metode *self-supervised learning* SimCLR. Model SimCLR dengan *backbone* ResNet-18 dilatih pada dataset Tiny ImageNet untuk mempelajari representasi fitur tanpa label. Dua modifikasi utama dieksperimenkan: penambahan augmentasi RandomSolarize untuk meningkatkan variasi data dan perampingan arsitektur *projection head* untuk menguji efisiensi model. Model dilatih selama 50 *epoch* menggunakan *optimizer* Adam dan *contrastive loss*. Hasil pelatihan dievaluasi melalui kurva *loss* yang menunjukkan konvergensi model yang stabil. Eksperimen ini membuktikan bahwa modifikasi yang diterapkan tetap mampu menghasilkan model yang belajar secara efektif, serta memberikan wawasan mengenai pengaruh komponen arsitektur terhadap performa SimCLR.**

Kata Kunci*—*Self-Supervised Learning, SimCLR, Contrastive Learning, Representasi Fitur, Augmentasi Data, ResNet-18

I. PENDAHULUAN

Klasifikasi citra merupakan salah satu domain fundamental dalam visi komputer (*computer vision*) dengan aplikasi yang luas di berbagai sektor. Kemajuan pesat dalam *deep learning*, khususnya penggunaan *Convolutional Neural Network* (CNN), telah menunjukkan efektivitas yang luar biasa dalam tugas pengenalan pola visual. Namun, performa model CNN konvensional sangat bergantung pada ketersediaan dataset berlabel dalam skala besar. Proses pelabelan data secara manual merupakan pekerjaan yang mahal, memakan waktu, dan tidak praktis untuk diterapkan pada jutaan data tidak berlabel yang tersedia. Untuk mengatasi tantangan ini, paradigma

Self-Supervised Learning (SSL) hadir sebagai solusi yang menjanjikan, memungkinkan model untuk mempelajari representasi fitur yang kaya dari data tanpa memerlukan anotasi manual.

Salah satu kerangka kerja SSL terkemuka adalah *A Simple Framework for Contrastive Learning of Visual Representations* (SimCLR), yang berbasis pada metode *contrastive learning*. SimCLR melatih sebuah *encoder* untuk memaksimalkan kesamaan (agreement) antara dua versi augmentasi dari citra yang sama (pasangan positif) dan meminimalkannya dengan citra lain (pasangan negatif). Kinerja SimCLR sangat dipengaruhi oleh komponen-komponen kuncinya, seperti strategi augmentasi data dan arsitektur *projection head*. Oleh karena itu, rumusan masalah dalam penelitian ini adalah bagaimana dampak dari modifikasi pada kedua komponen tersebut terhadap efektivitas proses pembelajaran representasi model SimCLR.

Tujuan dari penelitian ini adalah untuk: (1) mengimplementasikan dan melatih model SimCLR dengan *backbone* ResNet-18 pada dataset Tiny ImageNet; (2) melakukan eksperimen dengan memodifikasi pipeline augmentasi data dan arsitektur *projection head*; serta (3) menganalisis dampak dari modifikasi tersebut terhadap kinerja pelatihan model, yang dievaluasi berdasarkan konvergensi kurva *loss*.

II. TINJAUAN PUSTAKA

A. Self-Supervised Learning (SSL)

Self-Supervised Learning (SSL) merupakan salah satu cabang dari *machine learning* yang bertujuan untuk mempelajari representasi data dari dataset berskala besar tanpa memerlukan label yang dibuat oleh manusia. Metode ini bekerja dengan cara membuat tugas buatan, atau *pretext task*, di mana label untuk tugas tersebut dapat dihasilkan secara otomatis dari data itu sendiri. Model dilatih untuk menyelesaikan *pretext task* ini, dan dalam prosesnya, ia dipaksa untuk mempelajari fitur-fitur semantik yang esensial dari data. Representasi fitur yang telah dipelajari ini kemudian dapat ditransfer dan disesuaikan (*fine-tuned*) untuk menyelesaikan tugas-tugas hilir (*downstream tasks*) seperti klasifikasi atau deteksi objek dengan jumlah data berlabel yang jauh lebih sedikit.

B. Constarstive Learning dan SimCLR

The *Contrastive Learning* adalah pendekatan dominan dalam SSL visual yang bekerja dengan cara mengajarkan model tentang data mana yang "mirip" dan mana yang "berbeda". Tujuan utamanya adalah untuk mempelajari sebuah ruang representasi (*embedding space*) di mana sampel yang serupa (pasangan positif) ditarik agar berdekatan, sementara sampel yang tidak serupa (pasangan negatif) didorong agar berjauhan.

A Simple Framework for Contrastive Learning (SimCLR) adalah salah satu implementasi dari ide ini. Kerangka kerja SimCLR terdiri dari empat komponen utama:

*1) **Modul Augmentation Data:** Untuk setiap citra, dua transformasi augmentasi acak diterapkan untuk menghasilkan satu pasangan positif. Augmentasi yang kuat dan beragam adalah kunci dari keberhasilan SimCLR.*

*2) **Base Encoder:** Sebuah jaringan syaraf tiruan, seperti ResNet, digunakan untuk mengekstrak vektor representasi dari setiap citra yang telah di-augmentasi.*

*3) **Projection Head:** Sebuah Multi-Layer Perceptron (MLP) kecil ditambahkan setelah encoder untuk memetakan representasi ke ruang laten di mana contrastive loss dihitung.*

*4) **Contrastive Loss Function:** Fungsi loss (seperti NT-Xent) yang bertugas untuk memaksimalkan kesamaan antara pasangan positif dan meminimalkannya terhadap semua pasangan negatif lain dalam satu batch.*

C. *ResNet-18*

ResNet (*Residual Network*) adalah arsitektur *Convolutional Neural Network* (CNN) yang memperkenalkan konsep *residual learning* untuk memecahkan masalah *vanishing gradient* yang sering terjadi pada pelatihan jaringan yang sangat dalam. Inovasi utamanya adalah penggunaan "koneksi pintas" (

shortcut connections) yang memungkinkan gradien mengalir langsung melewati beberapa lapisan. Koneksi ini memungkinkan pembangunan model yang jauh lebih dalam daripada arsitektur sebelumnya sambil tetap mempertahankan kemudahan optimisasi. ResNet-50, salah satu varian populernya, memiliki 50 lapisan. Dalam penelitian ini, digunakan ResNet-18, varian yang lebih ringan dengan 18 lapisan, yang menawarkan keseimbangan yang baik antara kapasitas representasi dan efisiensi komputasi untuk dataset berskala menengah seperti Tiny ImageNet.

III. METODOLOGI PENELITIAN

A. *Dataset*

Penelitian ini menggunakan dataset **Tiny ImageNet**, yang merupakan himpunan bagian dari dataset ImageNet ILSVRC (ImageNet Large Scale Visual Recognition Challenge). Dataset ini terdiri dari 200 kelas citra dengan total 100.000 citra untuk pelatihan. Setiap citra memiliki resolusi 64x64 piksel, sehingga cocok untuk eksperimen yang membutuhkan waktu pelatihan yang lebih cepat dibandingkan menggunakan dataset ImageNet berskala penuh.

B. *Preprocessing dan Augmentasi Data*

Tahap *preprocessing* dan augmentasi data merupakan komponen krusial dalam kerangka kerja SimCLR. Setiap citra dalam *batch* pelatihan dilewatkan melalui serangkaian transformasi stokastik yang sama untuk menghasilkan dua "pandangan" (*views*) terkorélasi yang menjadi pasangan positif. Pipeline augmentasi yang diterapkan secara berurutan adalah sebagai berikut:

1) *RandomResizedCrop ke ukuran 64x64 piksel.*
2) *RandomHorizontalFlip dengan probabilitas 50%.*
3) *RandomSolarize dengan threshold 128 dan probabilitas 20% (modifikasi eksperimental).*
4) *RandomApply yang menerapkan ColorJitter (kecerahan, kontras, saturasi, rona) dengan probabilitas 80%.*
5) *RandomGrayscale dengan probabilitas 20%.*
6) *GaussianBlur dengan kernel acak.*
7) *Konversi citra ke format Tensor PyTorch.*

C. *Arsitektur Model*

Arsitektur model yang digunakan terdiri dari dua komponen utama, yaitu *base encoder* dan *projection head*.

- **Base Encoder:** Model **ResNet-18** digunakan sebagai *encoder* untuk mengekstraksi vektor fitur dari citra yang telah di-augmentasi. *Output* dari lapisan *average pooling* terakhir pada ResNet-18 menghasilkan vektor fitur dengan dimensi 512.
- **Projection Head:** Vektor fitur dari *encoder* kemudian dilewatkan ke sebuah *projection head*, yang merupakan *Multi-Layer Perceptron* (MLP) sederhana. Sesuai dengan modifikasi eksperimental, *projection head* ini dirancang dengan arsitektur yang lebih ramping, terdiri dari satu lapisan tersembunyi (*hidden layer*). Arsitekturnya adalah: satu lapisan Linear yang memetakan fitur dari 512 ke 256 dimensi, diikuti oleh aktivasi ReLU, dan diakhiri dengan lapisan Linear yang memetakan dari 256 ke ruang laten berdimensi 128.

D. *Konfigurasi Pelatihan*

Pelatihan model dilakukan dengan menggunakan konfigurasi sebagai berikut:

- **Optimizer:** Adam.
- **Loss Function:** InfoNCE Loss, yang diimplementasikan menggunakan CrossEntropyLoss untuk mengukur kesamaan kosinus (*cosine similarity*) antara representasi pasangan positif dan negatif.
- **Epoch:** Pelatihan dijalankan selama 50 *epoch*.
- **Batch Size:** 256.
- **Learning Rate:** 0.0003, dengan *scheduler* CosineAnnealingLR.
- **Temperature:** Parameter skalar (τ) pada *loss function* diatur ke 0.07.
- **Akselerasi:** Pelatihan menggunakan *mixed precision* (FP16) untuk mempercepat komputasi.

IV. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil dari proses pelatihan model SimCLR yang telah dijalankan sesuai dengan metodologi yang diuraikan pada bab sebelumnya. Analisis dilakukan terhadap metrik-metrik pelatihan untuk mengevaluasi kinerja model.

Pelatihan model dijalankan selama 50 *epoch*. Kinerja model selama proses pelatihan dipantau melalui nilai *loss function* pada setiap langkah iterasi. Kurva *training loss* yang dihasilkan disajikan pada Fig. 1.

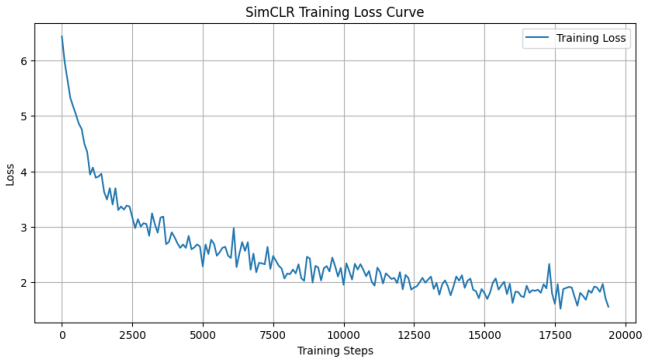
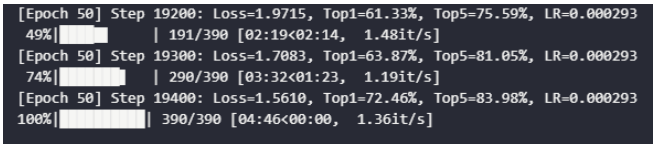


Fig. 1. Grafik Training Loss Model Selama 50 Epoch

Berdasarkan Fig. 1, dapat diamati bahwa model menunjukkan proses pembelajaran yang berhasil. Kurva *loss* menampilkan tren penurunan yang konsisten dan signifikan, terutama pada *epoch-epoch* awal, yang mengindikasikan bahwa model dengan cepat belajar untuk membedakan antara pasangan positif dan negatif. Menjelang akhir pelatihan, kurva mulai melandai dan menunjukkan tanda-tanda

konvergensi, di mana nilai *loss* menjadi lebih stabil di sekitar angka 1.5 hingga 2.0. Hal ini menandakan bahwa model telah mencapai titik optimal dalam mempelajari representasi fitur dari data latih.

Hasil akhir dari metrik pelatihan pada *epoch* ke-50 disajikan pada Fig. 2.



V. KESIMPULAN

Penelitian ini telah berhasil mengimplementasikan dan menganalisis dampak modifikasi arsitektur pada model *self-supervised learning* SimCLR yang dilatih pada dataset Tiny ImageNet. Dua modifikasi utama, yaitu penambahan augmentasi RandomSolarize dan perampingan arsitektur *projection head*, telah diuji. Berdasarkan hasil pelatihan selama 50 *epoch*, dapat disimpulkan bahwa model yang dimodifikasi menunjukkan performa pembelajaran yang baik. Hal ini dibuktikan dengan kurva *loss* yang konvergen secara stabil dan pencapaian *contrastive accuracy* Top1 yang tinggi, yang menandakan kualitas representasi fitur yang baik.

Temuan ini mengindikasikan bahwa arsitektur *projection head* yang lebih efisien secara komputasi (lebih ramping) tetap mampu mendukung proses *contrastive learning* secara efektif. Untuk penelitian di masa depan, disarankan beberapa pengembangan, antara lain: (1) melakukan evaluasi kuantitatif pada tugas hilir (*downstream task*) seperti klasifikasi untuk mengukur secara langsung kualitas representasi yang telah dipelajari; (2) melakukan studi ablasi dengan membandingkan performa model yang dimodifikasi terhadap model *baseline* tanpa modifikasi; dan (3) mengeksplorasi penggunaan arsitektur *backbone* yang berbeda, seperti EfficientNet, untuk potensi peningkatan efisiensi dan performa.

UCAPAN TERIMAKASIH

Saya berterima kasih kepada Universitas Darussalam Gontor atas dukungan dan fasilitas yang diberikan selama pelaksanaan penelitian ini. Ucapan terima kasih juga ditujukan kepada dosen pengampu mata kuliah yang telah memberikan bimbingan dan arahan yang sangat berharga dalam penyelesaian tugas dan laporan ini.

REFERENCES

[1] [1] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, “A Simple Framework for Contrastive Learning of Visual Representations,” in *Proc. Intl. Conf. on Machine Learning (ICML)*, 2020.

[2] [2] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR) 2016, pp. 770–778.

[3] [3] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A Large-Scale Hierarchical Image Database,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2009.