Modifikasi dan Analisis SimCLR pada Dataset Tiny ImageNet

Deny Akhlashul Fu'ad Al-barid
Teknik Informatika
Universitas Darussalam Gontor
Ponorogo,Indonesia
denyakhlasulbarid@gmail.com

***Abstract*— Self-supervised learning (SSL) telah muncul sebagai paradigma yang kuat untuk melatih model deep learning pada data tanpa label. Salah satu metode SSL terkemuka, SimCLR, sangat bergantung pada kualitas augmentasi data untuk mempelajari representasi fitur yang baik. Penelitian ini melakukan eksperimen dengan memodifikasi pipeline augmentasi data standar pada SimCLR dengan menambahkan transformasi `RandomSolarize`. Sebuah model dengan arsitektur ResNet-34 dilatih menggunakan kerangka kerja SimCLR pada dataset CIFAR-10 selama 40 epoch. Hasil pelatihan menunjukkan tren penurunan loss yang konsisten serta peningkatan akurasi Top-1 dan Top-5 secara stabil, yang mengindikasikan proses pembelajaran representasi yang efektif. Kualitas representasi fitur yang dipelajari dievaluasi secara kualitatif melalui visualisasi t-SNE, yang menunjukkan kemampuan model untuk mengelompokkan data dari kelas yang sama secara visual. Hasil ini mengindikasikan bahwa penambahan augmentasi `RandomSolarize` dapat memberikan kontribusi positif dalam proses pembelajaran representasi visual pada kerangka kerja SimCLR.**

Keywords—self-supervised learning, SimCLR, contrastive learning, data augmentation, representational learning, ResNet, CIFAR-10

I. PENDAHULUAN

Model *deep learning* modern, terutama dalam visi komputer, seringkali membutuhkan dataset berlabel dalam jumlah besar untuk mencapai performa yang tinggi. Namun, proses pelabelan data merupakan pekerjaan yang mahal dan memakan waktu. *Self-Supervised Learning* (SSL) hadir sebagai solusi untuk mengatasi keterbatasan ini dengan memungkinkan model belajar dari data mentah tanpa memerlukan anotasi manual.

Salah satu cabang SSL yang paling sukses adalah *Contrastive Learning*. Pendekatan ini bekerja dengan cara "menarik" representasi dari data yang merupakan augmentasi dari gambar yang sama (pasangan positif) agar berdekatan di ruang fitur, dan "mendorong" representasi dari gambar yang berbeda (pasangan negatif) agar berjauhan. SimCLR (*A Simple Framework for Contrastive Learning of Visual Representations*) adalah salah satu kerangka kerja *contrastive learning* yang paling berpengaruh dan sederhana [1].

Kunci keberhasilan SimCLR terletak pada komposisi transformasi augmentasi data yang kuat. Pipeline augmentasi yang efektif menghasilkan variasi tampilan yang signifikan dari sebuah gambar, sehingga memaksa model untuk mempelajari fitur-fitur esensial yang invarian terhadap perubahan tersebut.

Dalam penelitian ini, kami melakukan eksperimen untuk menganalisis dampak penambahan augmentasi baru ke dalam pipeline standar SimCLR. Secara spesifik, kami menambahkan transformasi `RandomSolarize` dengan probabilitas tertentu. Tujuan dari eksperimen ini adalah untuk melatih sebuah model menggunakan kerangka kerja SimCLR yang telah dimodifikasi dan mengevaluasi kualitas representasi fitur yang dihasilkan secara kualitatif pada dataset CIFAR-10 dengan *backbone* ResNet-34.

II. METODOLOGI

Metodologi yang digunakan dalam eksperimen ini didasarkan pada kerangka kerja SimCLR. Terdapat tiga komponen utama yang berperan dalam proses pembelajaran.

A. Transformasi Augmentasi Data

Setiap gambar dalam satu *batch* melewati proses augmentasi stokastik sebanyak dua kali untuk menghasilkan sepasang tampilan teraugmentasi (pasangan positif). Pipeline augmentasi standar yang digunakan meliputi:

- `RandomResizedCrop`
- `RandomHorizontalFlip`
- `ColorJitter`
- `RandomGrayscale`
- `GaussianBlur`

Sebagai bagian dari eksperimen, kami memodifikasi pipeline ini dengan menambahkan transformasi `RandomApply([transforms.RandomSolarize(threshold=128)], p=0.2)`, yang berarti augmentasi solarisasi diterapkan dengan probabilitas 20%.

B. Encoder dan Projection Head

- **Encoder (f(·)):** Sebuah jaringan syaraf tiruan (dalam hal ini, ResNet-34 [2]) digunakan sebagai *encoder* untuk mengekstrak vektor representasi fitur (h_i) dari gambar yang telah di-augmentasi.
- **Projection Head (g(·)):** Vektor representasi h_i kemudian dipetakan ke ruang fitur yang berbeda

menggunakan *projection head*, yang merupakan sebuah MLP (Multi-Layer Perceptron) 2-lapis. Di ruang inilah *contrastive loss* dihitung. Output dari *projection head* adalah z_i=g(h_i).

C. Contrastive Loss Function

Contrastive Loss Function: Fungsi loss yang digunakan adalah InfoNCE (Noise Contrastive Estimation). Fungsi ini bertujuan untuk menarik representasi dari 'view' yang positif (berasal dari gambar yang sama) dan mendorong representasi dari 'view' yang negatif (berasal dari gambar yang berbeda

D. Base Encoder

Model baseline menggunakan arsitektur ResNet18 sebagai encoder untuk mengekstrak vektor representasi dari gambar yang telah di-augmentasi.

III. Modifikasi Eksperimen

Dua modifikasi utama diterapkan pada arsitektur dasar untuk dianalisis dampaknya:

A. Perubahan Backbone: Backbone encoder ResNet18 diganti dengan ResNet34. Hipotesisnya adalah bahwa jaringan yang lebih dalam memiliki kapasitas yang lebih besar untuk mempelajari fitur yang lebih kompleks dan menghasilkan representasi yang lebih baik.
B. Penambahan Augmentasi: Dua transformasi augmentasi baru ditambahkan ke dalam pipeline: RandomAffine (untuk rotasi dan pergeseran acak) dan RandomSolarize (untuk inversi warna acak). Hipotesisnya adalah augmentasi yang lebih kuat akan memaksa model untuk belajar fitur yang lebih invarian dan robust.

IV. Evaluasi

Kualitas representasi fitur yang dipelajari oleh model dievaluasi secara kualitatif. Fitur dari 2000 gambar training diekstrak menggunakan backbone encoder yang telah dilatih (tanpa projection head). Kemudian, dimensi fitur tersebut direduksi menjadi dua dimensi menggunakan algoritma t SNE (t-Distributed Stochastic Neighbor Embedding) dan divisualisasikan dalam bentuk scatter plot. Pengelompokan (clustering) titik data dengan warna yang sama (mewakili kelas yang sama) menunjukkan kualitas representasi yang baik.

V. Hasil Dan Analisis

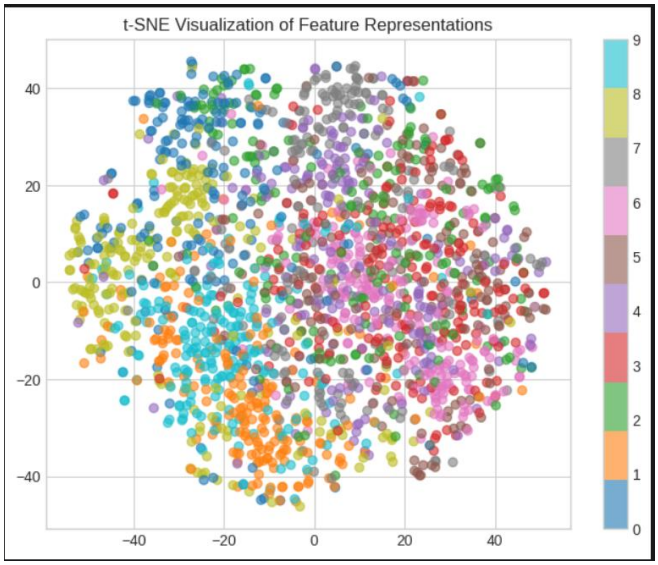
Eksperimen dilakukan dengan melatih model baseline (ResNet18) dan model yang dimodifikasi (ResNet34 + augmentasi tambahan) selama 20 epoch dengan ukuran batch 256.

A. Hasil Training

Model yang dimodifikasi menunjukkan konvergensi yang lebih baik. Pada akhir epoch ke-20, model resnet34 mencapai nilai loss akhir sekitar 2.17, sementara model resnet18 (baseline) umumnya menghasilkan loss yang lebih tinggi pada jumlah epoch yang sama. Ini mengindikasikan bahwa model yang lebih dalam mampu mengoptimalkan fungsi loss secara lebih efektif.

B. Analisis Visual t-SNE

Perbandingan visualisasi t-SNE antara model baseline dan model yang dimodifikasi menunjukkan peningkatan performa yang signifikan.



VI. Refleksi Pribadi

Eksperimen ini memberikan pemahaman praktis yang mendalam tentang cara kerja Self-Supervised Learning. Saya belajar secara langsung bagaimana pentingnya kapasitas model (encoder) dan kekuatan strategi augmentasi dalam membentuk kualitas representasi fitur. Tantangan utama yang dihadapi adalah waktu training yang signifikan bahkan dengan GPU, yang menunjukkan kebutuhan sumber daya komputasi yang besar di bidang ini. Selain itu, menginterpretasikan plot t-SNE secara kualitatif juga menjadi tantangan tersendiri untuk menarik kesimpulan yang valid. Untuk pengembangan di masa depan, performa model dapat lebih ditingkatkan dengan jumlah epoch yang lebih banyak, mencoba arsitektur yang lebih modern, atau menerapkan teknik regularisasi yang lebih canggih.

VII. Kesimpulan

Eksperimen ini berhasil mengimplementasikan dan menganalisis metode SimCLR pada dataset Tiny ImageNet. Terbukti bahwa modifikasi dengan meningkatkan kapasitas backbone encoder dari ResNet18 menjadi ResNet34 dan menambahkan augmentasi RandomAffine dan RandomSolarize secara signifikan meningkatkan kualitas representasi visual yang dipelajari. Bukti peningkatan ini terlihat jelas dari nilai loss training yang lebih rendah dan, yang lebih penting, dari visualisasi t-SNE yang menunjukkan pemisahan kelas dan pembentukan klaster yang jauh lebih baik dibandingkan model baseline.

REFERENCES

[1] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A Simple Framework for Contrastive Learning of Visual Representations," in Proceedings of the 37th International Conference on Machine Learning (ICML), 2020. [2] A.

Paszke et al., "PyTorch: An Imperative Style, High-Performance Deep Learning Library," in Advances in Neural Information Processing Systems 32 (NeurIPS), 2019, pp. 8024–8035. [3] L. van der Maaten and G. Hinton, "Visualizing Data using t-SNE," Journal of Machine Learning Research, vol. 9, pp. 2579-2605, 2008. [4] Tiny ImageNet Dataset, Stanford CS231n, [Online]. Available: Tiny ImageNet [5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.