

Analisis Eksperimen *Self-Supervised Learning* Menggunakan SimCLR pada Dataset Tiny ImageNet

NAME: RIZKY CAHYONO PUTRA

NIM: 442023611012

Informatics Engineering, science and
technology, University of Darussalam
Gontor

rizkycahyonoputra80@student.cs.unida
.gontor.ac.id

Abstrak—Laporan ini menyajikan hasil eksperimen *self-supervised learning* menggunakan metode SimCLR pada dataset Tiny ImageNet. Kami mengimplementasikan dan memodifikasi kode dasar SimCLR yang disediakan, dengan fokus pada pengujian dampak arsitektur backbone, struktur projection head, serta kombinasi batch size dan temperature. Tujuan utama dari eksperimen ini adalah untuk memahami bagaimana setiap komponen kunci dari SimCLR berkontribusi terhadap kualitas representasi visual yang dipelajari tanpa menggunakan label. Modifikasi dilakukan pada backbone dari ResNet18 ke ResNet50, arsitektur projection head dari MLP sederhana ke MLP yang lebih kompleks, dan nilai batch size serta temperature. Hasil eksperimen, yang mencakup kurva loss dan visualisasi representasi menggunakan t-SNE, menunjukkan bahwa... [Sertakan ringkasan hasil Anda di sini, misalnya: "penggunaan ResNet50 menghasilkan konvergensi yang lebih cepat dan kluster fitur yang lebih jelas."].

Kata Kunci: Self-supervised learning, SimCLR, contrastive learning, ResNet, Tiny ImageNet.

I. PENDAHULUAN

Self-supervised learning (SSL) telah menjadi pendekatan yang dominan untuk melatih model representasi visual tanpa mengandalkan label data yang mahal dan sulit didapat. Salah satu kerangka kerja SSL yang paling berpengaruh adalah **SimCLR (A Simple Framework for Contrastive Learning of Visual Representations)**. SimCLR bekerja dengan melatih model untuk memaksimalkan kesamaan (similarity) antara dua versi teraugmentasi dari gambar yang sama (pasangan positif), sambil secara bersamaan meminimalkan kesamaan dengan gambar lain dalam batch yang sama (pasangan negatif).

Tugas ini bertujuan untuk memberikan pemahaman praktis tentang SimCLR melalui eksperimen langsung. Kami akan mempelajari arsitektur inti dari SimCLR, yaitu modul augmentasi, backbone encoder, projection head, dan loss function (InfoNCE loss). Selanjutnya, kami akan melakukan serangkaian modifikasi pada kode dasar untuk menguji bagaimana perubahan pada arsitektur model dan hyperparameter memengaruhi proses pelatihan dan kualitas representasi fitur yang dihasilkan pada dataset Tiny ImageNet.

II. PENJELASAN KODE DASAR

A. Modul Augmentasi

Modul Augmentasi adalah kode yang Terletak di **ContrastiveLearningDataset.get.simclr_pipeline_transform** Modul ini menerapkan serangkaian augmentasi stokastik (seperti *RandomResizedCrop*, *RandomHorizontalFlip*, *ColorJitter*, *RandomGrayscale*, dan *GaussianBlur*) untuk menghasilkan dua "tampilan" yang berbeda dari satu gambar input. Kedua tampilan ini membentuk pasangan positif.

B. Backbone Encoder

Model ResNet (ResNet18 atau ResNet50) yang bertindak sebagai *feature extractor*. Model ini mengubah gambar input menjadi vektor representasi berdimensi tinggi.

C. Projection Head

Jaringan MLP yang terpasang di atas backbone encoder. Komponen ini memetakan representasi fitur ke ruang dimensi yang lebih rendah tempat *contrastive loss* dihitung.

D. InfoNCE Loss

Fungsi kerugian (*info_nce_loss*) yang merupakan inti dari SimCLR. Fungsi ini menggunakan kemiripan kosinus untuk menghitung seberapa dekat pasangan positif satu sama lain dan seberapa jauh mereka dari pasangan negatif.

III. MODIFIKASI YANG DILAKUKAN DAN ALASAN

Dalam eksperimen ini, kami melakukan tiga modifikasi utama pada kode dasar untuk mengkaji pengaruh setiap komponen:

A. Mengganti Backbone CNN: ResNet18 ke ResNet50

Alasan: ResNet50 memiliki arsitektur yang lebih dalam (50 lapisan) dibandingkan ResNet18 (18 lapisan). Dengan jumlah parameter yang lebih banyak, ResNet50 memiliki kapasitas representasi yang lebih besar. Kami berhipotesis bahwa model yang lebih kuat ini akan mampu mengekstrak fitur yang lebih diskriminatif, yang berpotensi menghasilkan konvergensi pelatihan yang lebih cepat dan representasi yang lebih optimal.

B. Mengubah Struktur Projection Head

Alasan: Kami mengubah arsitektur MLP pada *projection head* dari dua lapisan linear menjadi tiga lapisan linear yang lebih kompleks. Perubahan ini bertujuan untuk menguji apakah pemetaan non-linier yang lebih rumit dapat meningkatkan kualitas representasi di ruang proyeksi. *Projection head* memisahkan tugas ekstraksi fitur (backbone) dari tugas *loss* (proyeksi), dan kami ingin melihat apakah arsitektur yang lebih kompleks di bagian ini dapat mengoptimalkan pembelajaran.

C. Mengubah Batch Size dan Temperature

Alasan: Kami meningkatkan *batch size* dari 256 menjadi 512 dan mengubah nilai *temperature* dari 0.07 menjadi 0.1. Peningkatan *batch size* akan memberikan lebih banyak sampel negatif per iterasi, yang secara teori dapat membantu model membedakan antara lebih banyak gambar yang berbeda, sehingga menghasilkan representasi yang lebih robust. Sementara itu, nilai *temperature* yang lebih tinggi akan membuat *loss function* kurang sensitif terhadap perbedaan kecil dalam kemiripan, mendorong model untuk belajar dari sampel negatif yang "lebih sulit".

IV. HASIL EKSPERIMEN DAN ANALISIS

A. Analisis Kinerja Pelatihan Model SimCLR

Berdasarkan log yang dihasilkan selama 50 epoch, proses pelatihan model SimCLR dapat dianggap **berhasil dan menunjukkan progres yang sehat**. Hal ini dapat dilihat dari beberapa indikator kunci: tren penurunan *loss* (kerugian), peningkatan akurasi *top-1* dan *top-5*, serta stabilitas pelatihan secara keseluruhan.

B. Tren Penurunan Loss Yang Konsisten

Pada awal pelatihan (Epoch 1, Step 0), nilai *loss* dimulai sangat tinggi di angka **7.0970**. Nilai yang tinggi ini wajar terjadi karena bobot model diinisialisasi secara acak dan belum mampu membedakan antara pasangan gambar yang positif (augmentasi dari gambar yang sama) dan negatif (gambar yang berbeda). Seiring berjalannya waktu, nilai *loss* menunjukkan tren penurunan yang kuat dan konsisten. Pada pertengahan pelatihan (misalnya, Epoch 26, Step 5000), *loss* sudah turun drastis ke angka **2.7807**, dan pada akhir pelatihan (Epoch 50, Step 9700), *loss* mencapai **2.2665**. Penurunan yang stabil ini adalah indikator utama bahwa model berhasil mempelajari representasi yang bermakna. Model secara efektif belajar untuk menarik fitur dari pasangan positif agar lebih dekat satu sama lain di dalam ruang representasi, sambil mendorong pasangan negatif agar lebih jauh.

V. REFLEKSI PRIBADI

Selama pengerjaan tugas ini, saya memperoleh pemahaman yang mendalam tentang mekanisme *self-*

supervised learning dan bagaimana setiap komponen dalam kerangka kerja SimCLR bekerja sama. Tantangan terbesar yang saya hadapi adalah memahami peran setiap *hyperparameter* dan merumuskan modifikasi yang rasional dan beralasan. Misalnya, memahami trade-off antara *batch size* dan *temperature* memerlukan pemahaman yang lebih dalam tentang cara kerja InfoNCE loss.

Melalui eksperimen ini, saya belajar bahwa meskipun SimCLR adalah "kerangka kerja sederhana", performanya sangat sensitif terhadap pilihan augmentasi, arsitektur *backbone*, dan *hyperparameter*. Eksperimen ini juga menggarisbawahi pentingnya analisis hasil secara visual, seperti menggunakan t-SNE, untuk memvalidasi secara kualitatif kualitas representasi yang dipelajari. Untuk proyek ke depan, akan menarik untuk menguji metode augmentasi yang lebih canggih atau membandingkan SimCLR dengan kerangka kerja SSL lainnya seperti BYOL.

REFERENCES

The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use "Ref. [3]" or "reference [3]" except at the beginning of a sentence: "Reference [3] was the first ..."

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors' names; do not use "et al.". Papers that have not been published, even if they have been submitted for publication, should be cited as "unpublished" [4]. Papers that have been accepted for publication should be cited as "in press" [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

- [1] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *International Conference on Machine Learning*, 2020.
- [2] K. He, H. Fan, Y. Wu, and S. G. P. al., "Momentum contrast for unsupervised visual representation learning," in *The IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020.
- [3] J. Grill et al., "Bootstrap your own latent - a new approach to self-supervised learning," in *Advances in Neural Information Processing Systems*, 2020.
- [4] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *International Conference on Learning Representations*, 2015.