Self-Supervised Learning for Gestational Age Estimation from Low-Cost Doppler Ultrasound in Low-Resource Settings

Mohsen Motie-Shirazi[1]
mohsen.shirazi@dbmi.emory.edu

Edlyn Ramos[2]
edlynramos@wuqukawoq.org

Suchitra Chandrasekaran[3]
suchitra.chandrasekaran@emory.edu

Reza Sameni[1,4]
rsameni@dbmi.emory.edu

Peter Rohloff[2]
peter@wuqukawoq.org

Gari D. Clifford[1,4]
gari@gatech.edu

Nasim Katebi[1,2]
nkatebi@emory.edu

[1]Department of Biomedical Informatics, Emory University, Atlanta, GA, USA
[2]Center for Indigenous Health Research, Wuqu' Kawoq | Maya Health Alliance, Tecpán, Guatemala
[3]Department of Obstetrics and Gynecology, Emory University, Atlanta, GA, USA
[4]Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA, USA

Abstract

We present a self-supervised learning framework for gestational age (GA) estimation from low-cost, one-dimensional Doppler ultrasound recordings. A segment encoder was pretrained on unlabeled Doppler data using a hybrid approach that combines Simple Siamese Networks (SimSiam) and Variance-Invariance-Covariance Regularization (VICReg), and subsequently fine-tuned on clinically accurate GA labels. The resulting model achieved a mean absolute error of 1.19 weeks in 5-fold cross-validation and 0.87 weeks on an external test set, an improvement of 11% over fully supervised approaches and 29% over transfer learning approaches. Our approach leverages abundant unlabeled Doppler data to learn generalizable fetal signal representations, enabling accurate GA estimation in low-resource settings and yielding transferable embeddings for future maternal–fetal health applications.

1 Introduction

Limited access to timely and accurate antenatal care contributes substantially to maternal and neonatal mortality in low- and middle-income countries and remains a persistent challenge in rural or underserved areas of high-income countries as well [1–3]. Reliable gestational age (GA) estimation is essential for delivery planning, fetal growth monitoring, and identifying adverse outcomes such as fetal growth restriction (FGR) [4, 5]. In many rural settings, GA is estimated from maternal recall of the last menstrual period (LMP), which is prone to errors from recall bias, irregular cycles, and coarse granularity, often rounded to months, leading to misclassification of term births as preterm and potentially compromising prenatal care quality [6]. Obstetric ultrasound, the standard alternative, is often unavailable due to cost, maintenance needs, and limited trained personnel [7, 8].

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

Over the past decade, a low-cost ($\sim$ \$15), one-dimensional Doppler ultrasound (1D-DUS) system integrated with a smartphone-based mHealth platform has been deployed with traditional Indigenous midwives in rural Guatemala, contributing to a significant reduction in maternal and neonatal mortality in the region [9, 10]. These devices measure fetal cardiac activity via the Doppler frequency shift from moving cardiac structures. Large-scale community use has yielded thousands of real-world fetal recordings, enabling machine learning models for signal quality assessment [11], fetal heart rate estimation [12, 13], and hypertension screening [14], supporting prenatal decision-making in low-literacy, resource-limited environments.

This approach can also support GA estimation, as fetal cardiac activity changes during pregnancy with maturation of the autonomic nervous system [5, 15, 9, 14, 16], altering Doppler-measured heart rate patterns and variability. Prior deep learning work using a hierarchical attention network (HAN), with a segment encoder and a sequence encoder, achieved a mean absolute error (MAE) of $\sim$0.79 months ($\approx$3.4 weeks) [17], but was constrained by noisy LMP-based labels with month-level resolution.

To address this, we apply self-supervised learning (SSL), which leverages large unlabeled datasets to pretrain models before fine-tuning on smaller, accurately labeled cohorts. SSL has shown strong potential in biomedical time-series domains such as Electrocardiography (ECG) and Photoplethysmography (PPG) [18–20]. Here, the HAN segment encoder is pretrained on community-collected DUS segments using a hybrid SimSiam+VICReg objective to learn physiologically meaningful embeddings without GA labels, then fine-tuned on clinically accurate labels from early ultrasound biometry. To our knowledge, this is the first application of SSL to 1D-DUS–based GA estimation, aiming to reduce reliance on noisy labels and improve generalization for earlier FGR detection, optimized delivery planning, and better triage of high-risk pregnancies in low-resource settings.

2 Methods

2.1 Data Description

1D-DUS signals were recorded using the low-cost AngelSounds fetal Doppler probe, which detects fetal cardiac activity via the Doppler shift. The device was connected to a Google Pixel smartphone via an audio cable, and a custom Android application was developed for real-time acquisition, storage, and on-device analysis [8]. An illustration of the data collection setting is shown in Figure. 1(a).

Three datasets were used, collected in community and clinical settings in Guatemala and the USA, each serving a distinct role in model development (Table 1). The **Community-collected Guatemala** dataset, acquired by traditional Indigenous midwives during routine antenatal visits, contains 5,140 recordings (3,413 with sufficient good-quality recordings) with GA labels—when available—based on the mother's recalled LMP reported in one-month increments for months 5–9. The **Clinic-measured Guatemala** dataset was collected in Tecpán, Guatemala, using the same device. GA labels were based on the first clinical ultrasound, and subsequent Doppler recordings were obtained during home visits by trained nurses. After excluding FGR cases, 677 of 825 recordings (15–36 weeks GA) from 231 patients remained. The **Clinic-measured Georgia** dataset was collected at clinical research facilities at Emory University, Georgia, USA, also using the same device. GA labels were based on the first clinical ultrasound, with Doppler recordings acquired at later visits. After excluding FGR and abnormal growth cases, 24 of 67 recordings (16–32 weeks GA) from 25 pregnancies remained. All recordings were approximately 5 minutes in duration.

Table 1: Summary of datasets. "Total" refers to all collected $\sim$5-min recordings; "Good-quality" refers to the subset with sufficient good-quality 3.75 s segments for analysis.

Dataset	Total Recordings	Good-quality Recordings	Usage
Guatemala Community	5,140	3,413	SSL pretraining
Guatemala Clinic	825	677	Supervised development
Georgia Clinic	67	24	External test

2.2 Model Architecture and Training Strategies

Base Architecture. Gestational age estimation was performed using the hierarchical attention network (HAN) introduced by Katebi et al. [17], with the architecture illustrated in Fig. 1(b).

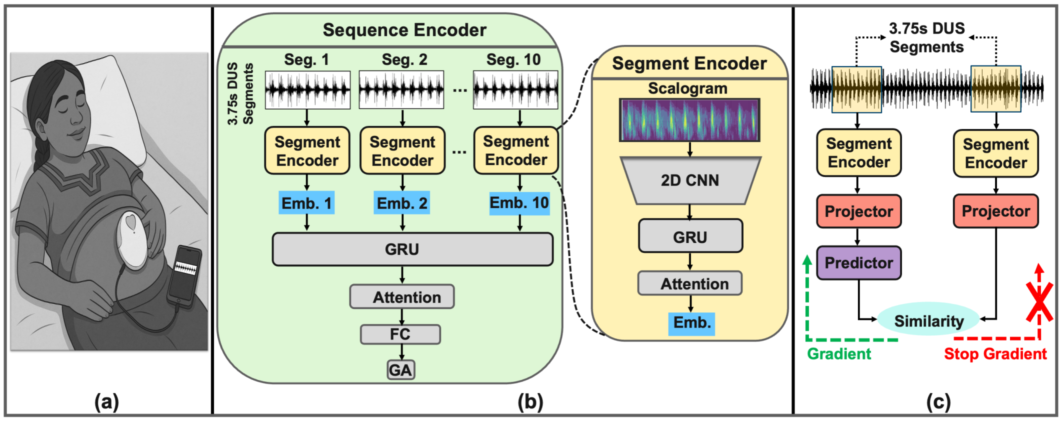

Figure 1: Overview of data collection and model architecture. (a) Illustration of 1D Doppler ultrasound (DUS) signal acquisition using a low-cost probe connected to a smartphone. (b) Hierarchical Attention Network for gestational age (GA) estimation from ten 3.75 s DUS segments per recording. (c) SimSiam framework used for self-supervised pretraining of the segment encoder, learning invariant representations from pairs of segments within the same recording.

Each Doppler recording was divided into non-overlapping good-quality 3.75 s segments, converted into scalograms, and processed by a segment encoder (2D CNN, GRU, attention) to produce 50-dimensional embeddings. A sequence encoder then aggregated embeddings from ten segments via GRU and attention layers, followed by a fully connected (FC) regression layer to output GA.

Self-Supervised Pretraining. In the proposed approach (Approach 1), the segment encoder, architecturally identical to the segment encoder of the HAN, was pretrained using self-supervised learning (SSL) on the Guatemala Community dataset. In total, 637,762 good-quality segments were available, identified using the signal quality assessment of Motie-Shirazi et al. [21]. Since all segments from the same recording correspond to the same GA, they are expected to share similar features in the embedding space. Therefore, positive pairs for SSL pretraining were generated by randomly selecting two segments from the same recording. We evaluated four self-supervised learning methods: Simple Siamese (SimSiam) [22], Bootstrap Your Own Latent (BYOL) [23], Variance–Invariance–Covariance Regularization (VICReg) [24], and a hybrid SimSiam+VICReg approach. After SSL pretraining with these methods, the encoder was integrated into the HAN sequence encoder and fine-tuned in a supervised manner on the Guatemala Clinic dataset.

The SimSiam framework, which yielded the best results when combined with VICReg (see Section 3.1), is shown in Figure. 1(c). It has two identical branches, each consisting of an encoder followed by a projection head implemented as a multi-layer perceptron (MLP). One branch also includes a predictor, also an MLP, and the loss is the negative cosine similarity between the predictor output and the projection from the other branch, with a stop-gradient applied to prevent collapse. Further details on SimSiam, VICReg, and their hybrid combination are provided in Section A.

Comparison Baselines. We compared the SSL-based approach (Approach 1) with two baselines: **Approach 2:** Transfer learning with a HAN pretrained on noisy LMP-based labels from the Guatemala Community dataset, with only the sequence encoder fine-tuned on the Guatemala Clinic dataset. **Approach 3:** HAN trained end-to-end from scratch on the Guatemala Clinic dataset.

Training and Evaluation. All approaches were trained and validated using 5-fold cross-validation, with the Georgia Clinic dataset serving as a held-out external test set. For the supervised stage, the mean absolute error (MAE) in weeks was used as the loss function. To reduce variability in segment selection, each recording was represented by 50 bootstrap samples of ten segments each; the final GA estimate was the median of these estimations.

3 Results

3.1 Self-Supervised Learning Method Comparison

Table 2 reports the 5-fold cross-validation MAE in weeks on the Guatemala Clinic dataset for models based on the segment encoders pretrained with each of the four SSL approaches. The hybrid SimSiam+VICReg achieved the best performance ($\mathbf{1.19 \pm 0.06}$ weeks), likely due to VICReg's variance and covariance regularization complementing SimSiam's invariance objective function.

Table 2: SSL method comparison on the Guatemala Clinic dataset (5-fold CV).

SSL Method	MAE (weeks)
BYOL	1.37 ± 0.05
VICReg	1.57 ± 0.08
SimSiam	1.29 ± 0.06
SimSiam+VICReg	$\mathbf{1.19 \pm 0.06}$

3.2 Comparison with Supervised Baselines

Table 3 shows the 5-fold CV results on the Guatemala Clinic dataset and the external test performance on the Georgia Clinic dataset. The SSL-pretrained SimSiam+VICReg approach achieved the lowest MAE in cross-validation and retained its best performance on the external test set.

Figure 2 shows estimated versus reference GA for this SSL-pretrained model on the Guatemala Clinic validation set (pooled CV folds) and the Georgia Clinic test set. Estimated values closely follow the identity line in both datasets, indicating strong agreement with the reference GA.

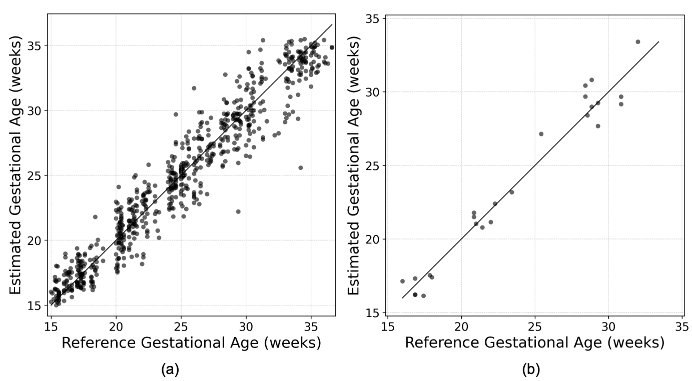

Figure 2: Estimated vs. reference GA for the SSL-pretrained model on (a) Guatemala Clinic and (b) Georgia Clinic datasets.

4 Discussion

Accurate GA estimation is critical for prenatal care, guiding delivery planning and early detection of complications. In this work, GA was referenced to the first ultrasound biometry performed during pregnancy before 19 weeks of gestation, the clinical gold standard with an accuracy of $\pm$5–10 days [25]. The proposed SSL-pretrained model (SimSiam+VICReg) achieved an MAE of 1.19 weeks ($\approx$8.3 days) in cross-validation and 0.87 weeks ($\approx$6.1 days) on the external test set, relying only on low-cost, one-dimensional Doppler signals. This performance is equivalent to the best estimates from accepted Doppler imaging approaches [25]. While an error of about a week can be clinically meaningful, it is low enough to be highly valuable in low-resource settings where no reliable GA estimate may otherwise be available, providing practical support for community-level care by traditional midwives and health workers.

Self-supervised learning enables the extraction of general physiological representations from Doppler signals without direct reliance on GA labels, resulting in embeddings that are broadly transferable to

Table 3: Performance comparison of SSL-pretrained and supervised models. MAE is reported as mean ± SD (weeks) for Guatemala Clinic (CV) and as the MAE value for Georgia Clinic (Test).

Model	Guatemala Clinic (CV)	Georgia Clinic (Test)
Approach 1: SSL (SimSiam+VICReg)	$\mathbf{1.19 \pm 0.06}$	$\mathbf{0.87}$
Approach 2: Noisy-label pretrain + fine-tune	1.66 ± 0.18	2.61
Approach 3: Fully supervised	1.34 ± 0.17	1.42

other maternal–fetal health applications. Future work will focus on leveraging these SSL-derived embeddings for fetal FGR classification, thereby extending the clinical scope beyond GA estimation.

5 Conclusion

Self-supervised pretraining of Doppler ultrasound embeddings enables GA estimation within about one week of accuracy, approaching clinical ultrasound standards while relying only on low-cost hardware and minimal infrastructure. This approach extends reliable pregnancy dating to low-resource environments and provides reusable representations for a broad range of fetal monitoring applications.

References

[1] World Health Organization. Newborn mortality. `https://www.who.int/news-room/fact-sheets/detail/newborn-mortality`, 2024.

[2] Joy E Lawn, Hannah Blencowe, Shefali Oza, Danzhen You, Anne CC Lee, Peter Waiswa, Marek Lalli, Zulfiqar Bhutta, Aluisio JD Barros, Parul Christian, et al. Every Newborn: progress, priorities, and potential beyond survival. *The Lancet*, 384(9938):189–205, 2014.

[3] Katy Backes Kozhimannil, Julia D Interrante, Carrie Henning-Smith, and Lindsay K Admon. Rural-urban differences in severe maternal morbidity and mortality in the US, 2007–15. *Health affairs*, 38(12):2077–2085, 2019.

[4] Russell Fung, José Villar, Ali Dashti, Leila Cheikh Ismail, Eleonora Staines-Urias, Eric O. Ohuma, Laurent J. Salomon, Cesar G. Victora, Fernando C. Barros, Ann Lambert, Maria Carvalho, Yasmin A. Jaffer, J Alison Noble, Michael G. Gravett, Manorama Purwar, Ruyan Pang, Enrico Bertino, Shama Munim, Aung Myat Min, Rose McGready, Shane A. Norris, Zulfiqar A. Bhutta, Stephen H. Kennedy, Aris T. Papageorghiou, Abbas Ourmazd, and the INTERGROWTH-21st Consortium. Achieving accurate estimates of fetal gestational age and personalised predictions of fetal growth based on data from an international prospective cohort study: a population-based machine learning study. *The Lancet Digital Health*, 2(7):e368–e375, 2020. doi: 10.1016/S2589-7500(20)30131-X.

[5] Camilo E Valderrama, Faezeh Marzbanrad, Rachel Hall-Clifford, Peter Rohloff, and Gari D Clifford. A proxy for detecting IUGR based on gestational age estimation in a Guatemalan rural population. *Frontiers in Artificial Intelligence*, 3:56, 2020.

[6] Joan T Price, Jennifer Winston, Bellington Vwalika, Stephen R Cole, Marie CD Stoner, Mwansa K Lubeya, Andrew Kumwenda, and Jeffrey SA Stringer. Quantifying bias between reported last menstrual period and ultrasonography estimates of gestational age in Lusaka, Zambia. *International Journal of Gynecology & Obstetrics*, 144(1):9–15, 2019.

[7] Gazala Mushtaq and K Veningston. A comprehensive review on fetal health surveillance using ultrasound imagery. *Discover Applied Sciences*, 7(5):478, 2025.

[8] Nasim Katebi, Mohammad Ahmad, Mohsen Motie-Shirazi, Daniel Phan, Ellen Kolesnikova, Sepideh Nikookar, Alireza Rafiei, Murali K. Korikana, Rachel Hall-Clifford, Esteban Castro, Rosibely Sut, Enma Coyote, Anahi Venzor Strader, Edlyn Ramos, Peter Rohloff, Reza Sameni, and Gari D. Clifford. Edge AI for Real-Time fetal assessment in rural Guatemala. In *Machine Learning for Healthcare (ML4H), Demo Track*, 2024.

[9] Lisa Stroux, Christopher W Redman, Antoniya Georgieva, Stephen J Payne, and Gari D Clifford. Doppler-based fetal heart rate analysis markers for the detection of early intrauterine growth restriction. *Acta obstetricia et gynecologica Scandinavica*, 96(11):1322–1329, 2017.

[10] Edlyn Ramos, Irma Piló Palax, Emily Serech Cuxil, Elsa Sebaquijay Iquic, Ana Canú Ajqui, Ann C Miller, Suchitra Chandrasekeran, Rachel Hall-Clifford, Reza Sameni, Nasim Katebi, et al. Mobil monitoring Doppler ultrasound (MoMDUS) study: protocol for a prospective, observational study investigating the use of artificial intelligence and low-cost Doppler ultrasound for the automated quantification of hypertension, pre-eclampsia and fetal growth restriction in rural Guatemala. *BMJ Open*, 14(9):e090503, 2024. doi: 10.1136/bmjopen-2024-090503.

[11] Camilo E Valderrama, Faezeh Marzbanrad, Lisa Stroux, and Gari D Clifford. Template-based quality assessment of the Doppler ultrasound signal for fetal monitoring. *Frontiers in physiology*, 8:511, 2017.

[12] Camilo E Valderrama, Lisa Stroux, Nasim Katebi, Elianna Paljug, Rachel Hall-Clifford, Peter Rohloff, Faezeh Marzbanrad, and Gari D Clifford. An open source autocorrelation-based method for fetal heart rate estimation from one-dimensional Doppler ultrasound. *Physiological measurement*, 40(2):025005, 2019.

[13] Alireza Rafiei, Mohsen Motie-Shirazi, Reza Sameni, Gari D. Clifford, and Nasim Katebi. Next-generation fetal heart monitoring: Leveraging neural sequential modeling for ultrasound analysis. *IEEE Transactions on Biomedical Engineering*, pages 1–12, 2025. doi: 10.1109/TBME.2025.3585461.

[14] Nasim Katebi, Gari D. Clifford, Reza Sameni, and Peter Rohloff. Deep sequence learning for assessing hypertension in pregnancy from Doppler signals. *medRxiv preprint*, 2022. doi: 10.1101/2022.01.26.22269921.

[15] Judith OEH Van Laar, Guy JJ Warmerdam, Kim MJ Verdurmen, Rik Vullings, Chris HL Peters, Saskia Houterman, Pieter FF Wijn, Peter Andriessen, Carola Van Pul, and S Guid Oei. Fetal heart rate variability during pregnancy, obtained from non-invasive electrocardiogram recordings. *Acta Obstetricia et Gynecologica Scandinavica*, 93(1):93–101, 2014. doi: 10.1111/aogs.12286.

[16] Fleur Zwanenburg, Monique RM Jongbloed, Nan Van Geloven, Arend DJ Ten Harkel, Jan MM Van Lith, and Monique C Haak. Assessment of human fetal cardiac autonomic nervous system development using color tissue doppler imaging. *Echocardiography*, 38(6):974–981, 2021. doi: 10.1111/echo.15094.

[17] Nasim Katebi, Reza Sameni, Peter Rohloff, and Gari D Clifford. Hierarchical attentive network for gestational age estimation in low-resource settings. *IEEE journal of biomedical and health informatics*, 27(5):2501–2511, 2023.

[18] Emadeldeen Eldele, Moamen Ragab, Zhenghua Chen, Min Wu, Chee Keong Kwoh, Xiaoli Li, and Cuntai Guan. Time-series representation learning via temporal and contextual contrasting. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 2352–2359. IJCAI, 2021. doi: 10.24963/ijcai.2021/324.

[19] Hubert Banville, Omar Chehab, Aapo Hyvärinen, Denis-Alexander Engemann, and Alexandre Gramfort. Uncovering the structure of clinical EEG signals with self-supervised learning. *Journal of Neural Engineering*, 18(4):051001, 2021. doi: 10.1088/1741-2552/abca18.

[20] Cheng Ding, Zhicheng Guo, Zhaoliang Chen, Randall J Lee, Cynthia Rudin, and Xiao Hu. SiamQuality: A ConvNet-based foundation model for imperfect physiological signals. *Physiological Measurement*, 45(8):085004, 2024. doi: 10.1088/1361-6579/ad6747.

[21] Mohsen Motie-Shirazi, Reza Sameni, Peter Rohloff, Nasim Katebi, and Gari D. Clifford. Point-of-care real-time signal quality for fetal Doppler ultrasound using a deep learning approach. In *Conference on Machine Learning for Health (ML4H)*, 2023.

[22] Xinlei Chen and Kaiming He. Exploring simple siamese representation learning. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15750–15758, 2021.

[23] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Guo, Mohammad Gheshlaghi Azar, et al. Bootstrap your own latent-a new approach to self-supervised learning. *Advances in neural information processing systems*, 33:21271–21284, 2020.

[24] Adrien Bardes, Jean Ponce, and Yann LeCun. VICReg: Variance-invariance-covariance regularization for self-supervised learning. *International Conference on Learning Representations (ICLR)*, 2022.

[25] American College of Obstetricians and Gynecologists. Methods for estimating the due date. *ACOG Committee Opinion No. 700*, May 2017. URL `https://www.acog.org/clinical/clinical-guidance/committee-opinion/articles/2017/05/methods-for-estimating-the-due-date`.

A Self-Supervised Learning Framework Details

A.1 SimSiam Framework

The SimSiam architecture used in this work is illustrated in Fig. 1(c) and described here for reference. Each branch consists of:

1. **Segment Encoder:** Adapted from the HAN segment encoder [17], taking 250×40 (time $\times$ frequency) scalograms as input. It contains three convolutional layers (32, 64, 128 channels), each followed by batch normalization, ReLU activation, max-pooling, and dropout. The output is reshaped and passed to a gated recurrent unit (GRU) and a hierarchical attention mechanism, yielding a 50-dimensional embedding.
2. **Projection Head:** A two-layer multi-layer perceptron (MLP) maps the 50-dimensional embedding to a 64-dimensional latent vector.

One branch additionally contains:

3. **Predictor:** A two-layer MLP mapping the 64-dimensional projection through a 32-dimensional hidden layer back to 64 dimensions.

Given two good-quality segments (x_1, x_2) from the same recording, the network outputs (p_1, z_1) and (p_2, z_2), where p denotes predictor outputs and z denotes projections. The SimSiam loss is:

$$\mathcal{L}_{\text{SimSiam}} = \frac{1}{2}\big[-\cos(p_1, \text{sg}(z_2)) - \cos(p_2, \text{sg}(z_1))\big]$$

where $\text{sg}(\cdot)$ is the stop-gradient operation to prevent collapse.

A.2 VICReg Regularization

VICReg (Variance–Invariance–Covariance Regularization) augments the SimSiam objective with additional variance and covariance regularization terms:

- **Invariance:** The SimSiam loss encourages similar embeddings for positive pairs.
- **Variance:** Applied to the projected embeddings z, this term ensures each embedding dimension has a batch standard deviation $\sigma_j \geq 1.0$:

$$\mathcal{L}_{\text{var}} = \frac{1}{d}\sum_{j=1}^{d} \max(0, 1 - \sigma_j), \quad \sigma_j = \sqrt{\text{Var}(z_{\cdot j}) + \epsilon}.$$

- **Covariance:** Penalizes redundancy by minimizing squared off-diagonal elements of the covariance matrix of z across the batch:

$$\mathcal{L}_{\text{cov}} = \frac{1}{d}\sum_{i \neq j} \text{cov}(z_i, z_j)^2$$

A.3 Combined Loss

The total SimSiam+VICReg loss is:

$$\mathcal{L} = \lambda_{\text{sim}}\mathcal{L}_{\text{SimSiam}} + \lambda_{\text{var}}\mathcal{L}_{\text{var}} + \lambda_{\text{cov}}\mathcal{L}_{\text{cov}}$$

where $(\lambda_{\text{sim}}, \lambda_{\text{var}}, \lambda_{\text{cov}}) = (1.0, 1.0, 0.01)$ in our experiments.