MICCAI

Learning Temporally Equivariance for Degenerative Disease Progression in OCT by Predicting Future Representations

Taha Emre, Arunava Chakravarty, Dmitrii Lachinov, Antoine Rivail,
Ursula Schmidt-Erfurth, and Hrvoje Bogunović

Department of Ophthalmology and Optometry, Medical University of Vienna, Austria
{taha.emre,hrvoje.bogunovic}@meduniwien.ac.at

Abstract. Contrastive pretraining provides robust representations by ensuring their invariance to different image transformations while simultaneously preventing representational collapse. Equivariant contrastive learning, on the other hand, provides representations sensitive to specific image transformations while remaining invariant to others. By introducing equivariance to time-induced transformations, such as disease-related anatomical changes in longitudinal imaging, the model can effectively capture such changes in the representation space. In this work, we propose a Time-equivariant Contrastive Learning (TC) method. First, an encoder embeds two unlabeled scans from different time points of the same patient into the representation space. Next, a temporal equivariance module is trained to predict the representation of a later visit based on the representation from one of the previous visits and the corresponding time interval with a novel regularization loss term while preserving the invariance property to irrelevant image transformations. On a large longitudinal dataset, our model clearly outperforms existing equivariant contrastive methods in predicting progression from intermediate age-related macular degeneration (AMD) to advanced wet-AMD within a specified time-window.

Keywords: Equivariance · Contrastive Learning · AMD · Prediction.

1 Introduction

The advent of contrastive self-supervised learning (SSL) methods showed that it is possible to learn informative and discriminative image representations by learning invariances to image transformations that do not alter their semantics. The invariance is achieved by increasing the representational similarity between the original image and its transformed versions in the representation space, while pushing all pairs apart as negatives using a Siamese network [1]. However, sensitivity (*equivariance*) to certain transformations may be crucial for specific downstream tasks, such as rotation degree information in histopathology slides [21]. Introducing equivariance to time-induced transformations, such as anatomical changes due to degenerative disease progression in longitudinal

medical images could be pivotal in downstream tasks such as predicting the future risk of conversion of a patient from an early to a late-stage.

Retinal degenerative diseases such as Age-related Macular Degeneration (AMD) are frequently identified through non-invasive optical coherence tomography (OCT) scans, made up of sequential 2D cross-sectional slices (B-scans). As a leading cause of vision loss and blindness [3], AMD leads to irreversible tissue damage, implying that the severity of the disease can be represented as a monotonically non-decreasing function over time. However, the speed of disease progression and the associated anatomical changes vary widely between patients. Also, the longitudinal nature of the AMD datasets makes them costly to label by clinicians. Consequently, a Time-Equivariant SSL stage could be crucial in identifying the patients at a higher risk of progression to an advanced disease stage from a single visit to provide timely treatment and personalized management.

Existing *Equivariant SSL* methods aim to learn such sensitivity, unlike contrastive methods that seek total invariance. Indeed, invariance to certain transformations is not always desirable [22]. This sensitivity can be achieved with pretext SSL, by learning to recover the parameter of the target transformation from the representations. Image rotation prediction [11], video clip ordering [23] and distinguishing close and far clips [13] for temporal equivariance can be given as examples. In medical imaging, the time difference [19] or ordering two visits (scans) of a patient [16] is commonly used as a pretraining step for temporal tasks such as disease forecasting. In [27] they decoupled the effect of aging from disease-related changes in brain MRI scans by carefully curating a temporal dataset consisting of old subjects without any underlying diseases. Then, they learned an aging trajectory in the representation space shared by all patients. Although these methods learn sensitivity to the transformation, they lack the desired invariance to image perturbations.

To address these issues, *Equivariant contrastive methods* have been explored to learn equivariance for specific transformations while ensuring invariance for the others. Unlike methods with specific architecture [6] for achieving equivariance, equivariant contrastive learning methods are mostly architecture agnostic. Dangovski et al. [7] combined transformation parameter prediction with a contrastive loss for the transformation invariance, requiring four additional encoders for the rotation prediction task [11], increasing the computational overhead substantially (*ESSL*). Similarly, in [14], they proposed an additional temporal pretext task, such as frame ordering, to the contrastive task for videos. Emre et al. [9] introduced *TINC* for OCT scans where they exploited the time difference between visits by applying a similarity insensitivity increasing with the difference. In [17], *AugSelf* is proposed to enforce equivariance on the representation space by predicting transformation parameters between input pairs, then projected the learned representations into an invariant space for the contrastive task. However, transformation parameter recovery does not guarantee explicit equivariance, because it's possible for transformation parameters to be confined to a single dimension of the representation, leaving the rest completely invariant. In addition, parameter prediction models do not learn any explicit

function for directly transforming the representations. To address these limitations, Devillers et al. [8] proposed utilizing two different projections of the representation space; one for the invariance, and the second for the equivariance with a trainable module allowing direct manipulation of the projections with the image transformation parameter (*EquiMod*). Since invariance is a trivial solution for the equivariance where the equivariant representation transformation collapses to identity, they introduced an additional contrastive loss for the equivariance branch at the cost of an increased computational overhead. These methods aim to achieve equivariance using transformation parameter recovery, additional contrastive tasks, projection spaces, or pre-defined image transformation (e.g. rotation matrix) [10].

We propose to learn time-sensitive/equivariant representations directly on the *representation space* and learn to propagate representations in time, using longitudinal medical datasets in a contrastive SSL setting. Given that degenerative diseases either progress or stagnate over time, we have modeled disease progression in the representation space as an additive transformation parameterized by the time difference of a patient's visits, enabling direct temporal transformation of representations. We adapted a regularization loss on the calculated additive term to prevent the network from ignoring the time difference and producing identity transformation. Finally, we evaluated the learned representations for AMD progression prediction across multiple time-windows from a single visit scan. In summary, **our contributions** are (a) constructing contrastive pairs from different visits of a patient, (b) training a novel equivariance module to propagate representation in the future without accessing future visit scan, (c) introducing a novel regularization loss to prevent invariance to time.

2   Methodology

Our approach learns temporal equivariance in representation space, followed by a projection to invariant space within a contrastive SSL framework (Fig. 1). We propose a novel loss term specifically tailored for the equivariance (Eq. 5), utilizing a longitudinal retinal OCT dataset (Sec. 3.1) collected across multiple patient visits. First, we define the term time equivariant representation. Then, we detail our approach to learn the equivariance.

Time-Equivariant Representations The disease progression in a patient is routinely monitored using a series of medical scans $\boldsymbol{x} \in \mathbb{I} \subseteq \mathbb{R}^{H \times W}$, and $\boldsymbol{x}_t$ denotes the scan acquired at a time t sampled from $\mathbb{T} = \{t \in \mathbb{N}_0; t \leq b\}$, b being the last visit date. Let $\mu : \mathbb{I} \times \mathbb{T} \to \mathbb{I}$ represent a transformation in the image space for all possible time differences, which captures the complex anatomical changes between $\boldsymbol{x}_t$ and $\boldsymbol{x}_{t+\Delta t}$ due to the disease progression (and/or normal aging), where $t + \Delta t \in \mathbb{T}$. Although the actual transformation μ is difficult to model in the image space, the result of the transformation is available as $\boldsymbol{x}_{t+\Delta t}$ for some specific Δt, from available visits of the patient, such that $\boldsymbol{x}_{t+\Delta t} := \mu(\boldsymbol{x}_t, \Delta t)$. Let $f_\theta : \mathbb{I} \to \mathbb{R}^D$ with learnable parameters θ be a deep
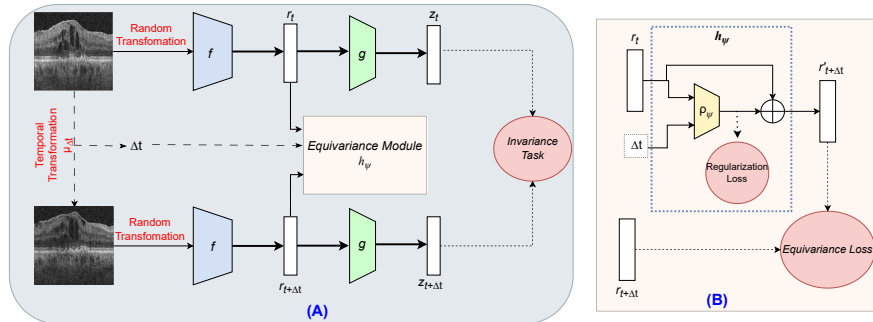
Fig. 1. An overview of **TC** architecture. **(A)** Representations of two scans acquired at two time points and the time difference are fed into h_ψ for equivariance, then projected for the invariance loss. **(B)** Equivariance module with an additive displacement map.

learning based encoder, which can embed $\boldsymbol{x}_t$ as a representation $r_t := f_\theta(\boldsymbol{x}_t)$ and $\boldsymbol{x}_{t+\Delta t}$ as $r_{t+\Delta t} := f_\theta(\boldsymbol{x}_{t+\Delta t})$. As $\mu(\cdot, \Delta t)$ transforms $\boldsymbol{x}_t$ to $\boldsymbol{x}_{t+\Delta t}$ in the image space, a corresponding transformation $h(\cdot, \Delta t)$ operates in the representation space such that $r_{t+\Delta t} = h(r_t, \Delta t)$. If f_θ is equivariant in time, the **equivariant relation** is defined as:

$$\exists h : \forall \Delta t \in \mathbb{T}, \quad f_\theta(\mu(\boldsymbol{x}_t, \Delta t)) = f_\theta(\boldsymbol{x}_{t+\Delta t}) \approx h(f_\theta(\boldsymbol{x}_t), \Delta t) \tag{1}$$

An undesired trivial solution for Eq. 1 would be *time-invariance* of f_θ such that $f_\theta(\boldsymbol{x}_t) = f_\theta(\boldsymbol{x}_{t+\Delta t})$ with disease progression being ignored, in turn making $h(\cdot, \Delta t)$ an identity mapping. This implies that f_θ learns only patient-specific features staying constant with respect to time while ignoring changes due to the progression. The transformation $h(\cdot, \Delta t)$ should be able to capture the future changes in the scans specific to each patient, and we propose to approximate it with a deep neural network that can model complex changes.

Time-Equivariant Contrastive Learning (TC) Our model extends the existing contrastive methods with the introduction of a temporal equivariance property. The input pair to the Siamese network is created from two different visits of a patient and both are transformed with contrastive augmentations (Fig.1.A). Recent equivariant contrastive methods with a learnable transformation [8,10] on features, relied on two projection spaces, one for equivariance and the other for invariance, preventing them from directly steering the representations. Instead, we propose to enforce equivariance directly on the representation space, and then to learn the invariance on a *projection* space [15] (Fig.1.A). This allows for the preservation of the time-sensitive information in the representation space while enforcing invariance in the projection space. A logical order since invariance is a trivial solution of equivariance. It also simplifies the

network architecture and computational overhead, and facilitates transferring the encoder weights as a backbone.

We introduce an equivariance module h_ψ containing the learned transformation (**predictor**) with learnable parameters ψ, which takes the concatenation of representation r_t and time difference Δt, normalized between 0-1, as input, and transforms r_t to $r'_{t+\Delta t}$ to satisfy the equivariance property defined in Eq. 1. The predictor enables the model to generate future visits' representations by forwarding available scans in time to be used in predictive tasks. The equivariance loss [8,10] for h_ψ is defined as:

$$\ell_{\text{equiv}} = \|f_\theta(\boldsymbol{x}_{t+\Delta t}) - h_\psi(f_\theta(\boldsymbol{x}_t), \Delta t)\|_2^2 = \|r_{t+\Delta t} - h_\psi(r_t, \Delta t)\|_2^2 \qquad (2)$$

As it is highlighted in [8,10], the contrastive loss imposes invariance to the image transformations including time-related changes. The learned invariance results in a trivial solution for the equivariance loss where h_ψ collapses to an identity mapping, ignoring Δt. The aforementioned methods rely on another set of computationally heavy contrastive task on $r'_{t+\Delta t}$ to prevent collapse. For an MLP h_ψ, the collapse manifests as $\frac{\partial h_\psi}{\partial t}$ becoming 0. To avoid an additional contrastive task or a regularization term on the gradients to prevent h_ψ from collapsing, we reparametrized h_ψ as an additive **displacement map** (DM):

$$r'_{t+\Delta t} = h_\psi(r_t, \Delta t) = r_t \oplus \rho_\psi(r_t, \Delta t) \qquad (3)$$

Now, the equivariant transformation h_ψ has two parts: an MLP ρ_ψ for predicting DM as an additive term and *direct sum* of the DM with the representation (Fig. 1.B). In Eq. 3, the collapse condition is much easier to regularize since it is the zero norm for the predicted DM. The ideal regularization should prevent DM norm becoming zero, while not encouraging high norm values. The representation space should respond to the time change smoothly without predicting large DMs easily. We adapted a regularization loss term for preventing collapse from pair-wise loss of RankNet [4]. The original loss function aims to increase the probability of ranking two embeddings correctly. In our case, the ranking is always one-sided, because the norm cannot take negative values. Thus, the final DM regularization term becomes:

$$\ell_{\text{reg}} = \log\left(1 + \exp(-\|\rho_\psi(r_t, \Delta t)\|_2)\right) \qquad (4)$$

The ℓ_{reg} penalizes null displacements without encouraging very high norms.

For the invariant contrastive loss, we used VICReg [2]. It is calculated using projections z_t and $z_{t+\Delta t}$, obtained by projecting representations r_t and $r_{t+\Delta t}$ with an MLP projector g_γ. As a *non-contrastive* method, it does not require very large batch sizes. It has 3 loss terms; an ℓ_2 term for increasing similarity within pair, a covariance term to prevent *dimensional collapse* [26,15] and a variance term for implicitly pushing the negatives apart which is calculated along a batch of projections, where each pair could have different Δt values.

Thus, the final training loss with a contrastive loss ℓ_{contr} is:

$$\mathcal{L}_{\text{TC}} = \ell_{\text{contr}} + \beta \cdot (\ell_{\text{equiv}} + \upsilon \cdot \ell_{\text{reg}}) \qquad (5)$$
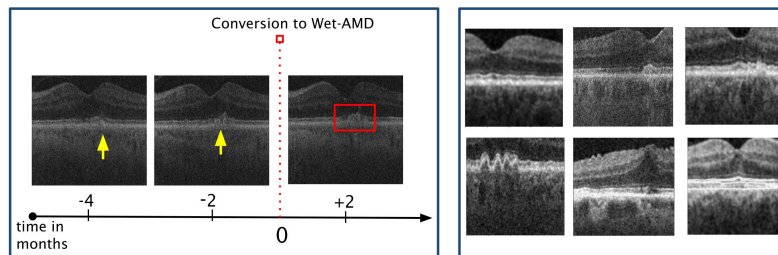
Fig. 2. Left: OCT scans with conversion to wet-AMD over time. **Right**: Contrastive augmentations examples.

3   Experiments and Results

We tested **TC** against three main equivariant contrastive methods: ESSL [7], AugSelf [17] and EquiMod [8]. ImageNet pretrained ResNet50, VICReg [2] and TINC [9] were used as baselines. During pretraining each batch contains only a single image pair from a patient to avoid pushing apart the intra-patient scans with the contrastive loss. Clinically, only future representations matter for conversion risk assessment. Therefore, the equivariant models were pre-trained only with image pairs with a positive Δt.

3.1   Experimental Setup

Dataset   The SSL stage and the downstream evaluation were conducted on the HARBOR dataset (Fig. 2 left), comprising OCT scans from 1,096 patients' *fellow-eyes*[1] monitored for wet-AMD onset, with monthly follow-ups over 24 months. Intermediate stage visits (463 eyes - 10,108 scans) are used for predicting the onset of late, wet-AMD stage within 6 and 12 months [24,25,20] (117 converter eyes) as a forecasting task. The dataset was split with the eye level stratification into 4-fold cross-validation (80%) for hyper-parameter optimization and a hold-out test set (20%) for reporting the performance. The remaining 540 eyes (12,506 scans) are only used in contrastive pretraining step. We used the fovea centered **2D B-scan**, which is known to be most representative of the AMD state [18], resized to 224×224.

We followed the work of [9,12] for contrastive image transformations specific to B-scans (Fig. 2 right). During pretraining, for each input pair, Δt is **randomly** sampled from an interval of 1 to 12 months and normalized between $0 - 1$. The rationale behind the interval is that the set of all pair permutations for all possible time differences is very large (276 per patient) for a plausible training time. By limiting Δt, training h_ψ becomes much more convenient, because it only needs to produce DMs for a constrained but practically meaningful interval.

---

[1] The other eye that is not part of the clinical trial

Table 1. Linear evaluation results for wet-AMD conversion prediction within two time-windows: 6 and 12 months

Model	6-months			12-months		
	AUROC ↑	PRAUC ↑	BAcc ↑	AUROC ↑	PRAUC ↑	BAcc ↑
ImageNet	0.714 ± 0.018	0.106 ± 0.007	0.648 ± 0.016	0.692 ± 0.014	0.167 ± 0.009	0.616 ± 0.045
VICReg[2]	0.700 ± 0.029	0.137 ± 0.031	0.650 ± 0.021	0.671 ± 0.033	0.158 ± 0.019	0.622 ± 0.023
TINC[9]	0.729 ± 0.017	0.151 ± 0.019	0.652 ± 0.007	0.688 ± 0.060	0.210 ± 0.026	0.629 ± 0.055
ESSL[7]	0.718 ± 0.033	0.157 ± 0.034	0.649 ± 0.028	0.699 ± 0.041	0.187 ± 0.039	0.637 ± 0.032
AugSelf[17]	0.752 ± 0.031	$\mathbf{0.190 \pm 0.027}$	0.670 ± 0.029	0.735 ± 0.049	0.245 ± 0.040	0.651 ± 0.057
EquiMod[8]	0.734 ± 0.035	0.153 ± 0.038	0.652 ± 0.048	0.718 ± 0.059	0.226 ± 0.055	0.657 ± 0.031
TC w/o DM	0.761 ± 0.032	0.163 ± 0.027	0.658 ± 0.034	0.748 ± 0.033	0.220 ± 0.026	0.653 ± 0.031
TC w/o ℓ_{reg}	0.732 ± 0.021	0.137 ± 0.024	0.661 ± 0.009	0.721 ± 0.035	0.195 ± 0.025	0.655 ± 0.028
TC	$\mathbf{0.769 \pm 0.027}$	0.185 ± 0.032	$\mathbf{0.677 \pm 0.026}$	$\mathbf{0.752 \pm 0.034}$	$\mathbf{0.265 \pm 0.043}$	$\mathbf{0.671 \pm 0.032}$
TC_{syn}	0.784 ± 0.022	0.188 ± 0.036	0.683 ± 0.024	0.775 ± 0.030	0.304 ± 0.052	0.695 ± 0.023

Implementation details A ResNet50 with a representation dimension of 2048 is used as the encoder f_θ. Following the common practice of projecting representations to higher dimensions [2,26], the projector g_γ is a three-layer MLP with hidden dimensions of 4096. DM predictor ρ_ψ of TC is implemented as a two-layer MLP with an input size of $(2048+1)$ for $[r_t, \Delta t]$, and it outputs a DM with the same dimension as the representations. In EquiMod, ESSL and AugSelf, we followed the original works for the predictor architecture.

All models were pretrained for 300 epochs with the AdamW optimizer with a learning rate of 5E-4 and a batch size of 128. A cosine scheduler with a warm-up was used for decaying learning rates. The same weight decay of 1E-6 was applied to all pretraining setups. β and υ of $\mathcal{L}_{\mathrm{TC}}$ were set to 1 and 0.5, respectively.

3.2 Results

Downstream prediction performance After pretraining, we evaluated the extracted representations with a linear classifier on the downstream tasks (Sec. 3.1), as it is common practice in contrastive SSL [5]. A scan from a single time-point is used for future conversion prediction. The linear classifier is trained for 50 epochs with Adam optimizer and a learning rate of 1E-4. We reported the results (Table 1) with AUROC, PRAUC and Balanced Accuracy (BAcc) due to a large class imbalance (6-month 1:20, 12-month 1:10). Accordingly, the random prediction PRAUC for 6-month is 0.05, and for 12-month is 0.11.

Our TC model surpassed all the other ones in AUROC and BAcc in both tasks. In terms of 12-month PRAUC, TC outperformed all while achieving competitive 6-month PRAUC performance. Interestingly, AugSelf achieved better results against the comparison methods, which can be attributed to its enforcement of the equivariance in the representation space rather than the projection space. A decline in 12-month prediction accuracy was noted for all the models, likely due to the extended time frame and the fact that the predictive biomarkers were more subtle. Finally, the performance gains of equivariant contrastive methods over the others (Table 1 line 1-3) underscore the importance of including temporal equivariance in disease progression prediction.

Evaluation of the equivariance module Each scan representation is synthetically propagated in time by 6 months using equivariance module h_ψ without accessing the future scans. Then they were combined with the initial representations by averaging. Their predictive capability was assessed using the same linear evaluation, with the improved performances in Table 1 - TC_{syn} demonstrating the equivariance module's ability to discern significant disease progression from a single scan. Notably, this performance increase was more distinct in the 12-month task, where the biomarkers are less distinguishable due to the longer conversion time-frame. Similar improvement is not possible with the other equivariant contrastive methods since they cannot manipulate the representation space.

Ablation study We investigated the impact of the explicit DM prediction (Table 1 - TC w/o *DM*) and the regularization loss term (Table 1 - TC w/o ℓ_{reg}) components of TC by removing each at the time. In both cases, the downstream task performance degraded considerably underscoring the importance of both components. When compared against the non-equivariant methods, the ablated models still perform better highlighting the importance of time component. When TC w/o *DM* compared against TC w/o ℓ_{reg}, it performs considerably better. This can be attributed to TC w/o *DM* directly predicting the future representation without relying on an unregularized DM. This highlights the significance of ℓ_{reg} to predict meaningful DMs.

Limitations The main limitation of TC is that the displacement map and the Siamese pair forming rely on the assumption of irreversible disease progression. Even though the most degenerative diseases fall into this category, it limits the general use of the future prediction module for the other temporal tasks.

4   Conclusion

Predicting late-stage disease onset is challenging due to the varying progression speeds among patients and subtle prognostic biomarkers. In this study, we proposed exploiting temporal information in unlabeled longitudinal OCT datasets. We introduced Time-equivariant Contrastive (TC) SSL with a learnable equivariance module for directly propagating image representations in time. The module retains characteristics of disease progression, such as irreversible tissue loss or degeneration, over time. A novel regularization loss was introduced to avoid the trivial solution for the equivariance induced by the contrastive loss with minimal computational overhead, unlike other equivariant models. We tested TC for disease progression on a longitudinal dataset consisting of eye scans with different stages of AMD. The TC pretraining consistently outperformed other equivariant contrastive methods as well as non-equivariant ones. Results highlight the importance of temporal sensitivity in representations for accurately assessing a patient's risk of conversion to a late disease stage. Furthermore, we found that integrating the synthetically propagated representations with the training data enhanced the conversion prediction performance, illustrating not only that TC

captured relevant longitudinal information but also provided a functional prediction module, demonstrating its ability to generate a proper time-equivariant representation space.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Azizi, S., Mustafa, B., Ryan, F., Beaver, Z., Freyberg, J., Deaton, J., Loh, A., Karthikesalingam, A., Kornblith, S., Chen, T., et al.: Big self-supervised models advance medical image classification. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 3478–3488 (2021)
2. Bardes, A., Ponce, J., LeCun, Y.: Vicreg: Variance-invariance-covariance regularization for self-supervised learning. In: International Conference on Learning Representations (2022)
3. Bressler, N.M.: Age-Related Macular Degeneration Is the Leading Cause of Blindness . . . JAMA **291**(15), 1900–1901 (04 2004). https://doi.org/10.1001/jama.291. 15.1900, https://doi.org/10.1001/jama.291.15.1900
4. Burges, C., Shaked, T., Renshaw, E., Lazier, A., Deeds, M., Hamilton, N., Hullender, G.: Learning to rank using gradient descent. In: International Conference on Machine Learning. pp. 89– 96 (2005)
5. Chen, X., Xie, S., He, K.: An empirical study of training self-supervised vision transformers. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 9620–9629 (2021)
6. Cohen, T.S., Welling, M.: Group equivariant convolutional networks. In: International Conference on Machine Learning. PMLR, JMLR.org (2016)
7. Dangovski, R., Jing, L., Loh, C., Han, S., Srivastava, A., Cheung, B., Agrawal, P., Soljacic, M.: Equivariant self-supervised learning: Encouraging equivariance in representations. In: International Conference on Learning Representations (2022), https://openreview.net/forum?id=gKLAAfiytI
8. Devillers, A., Lefort, M.: Equimod: An equivariance module to improve visual instance discrimination. In: International Conference on Learning Representations (2023), https://openreview.net/forum?id=eDLwjKmtYFt
9. Emre, T., Chakravarty, A., Rivail, A., Riedl, S., Schmidt-Erfurth, U., Bogunović, H.: Tinc: Temporally informed non-contrastive learning for disease progression modeling in retinal oct volumes. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 625–634. Springer (2022)
10. Garrido, Q., Najman, L., Lecun, Y.: Self-supervised learning of split invariant equivariant representations. In: International Conference on Machine Learning. PMLR (2023)
11. Gidaris, S., Singh, P., Komodakis, N.: Unsupervised representation learning by predicting image rotations. In: International Conference on Learning Representations (2018)

12. Holland, R., et al.: Metadata-enhanced contrastive learning from retinal optical coherence tomography images. CoRR **abs/2208.02529** (2022)
13. Jayaraman, D., Grauman, K.: Slow and steady feature analysis: higher order temporal coherence in video. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3852–3861 (2016)
14. Jenni, S., Jin, H.: Time-equivariant contrastive video representation learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9970–9980 (2021)
15. Jing, L., Vincent, P., LeCun, Y., Tian, Y.: Understanding dimensional collapse in contrastive self-supervised learning. In: International Conference on Learning Representations (2022), https://openreview.net/forum?id=YevsQ05DEN7
16. Kim, H., Sabuncu, M.R.: Learning to compare longitudinal images. In: Medical Imaging with Deep Learning (2023)
17. Lee, H., Lee, K., Lee, K., Lee, H., Shin, J.: Improving transferability of representations via augmentation-aware self-supervision. In: Beygelzimer, A., Dauphin, Y., Liang, P., Vaughan, J.W. (eds.) Advances in Neural Information Processing Systems (2021), https://openreview.net/forum?id=U34rQjnImpM
18. Lin, A.C., Lee, C.S., Blazes, M., Lee, A.Y., Gorin, M.B.: Assessing the clinical utility of expanded macular octs using machine learning. Translational vision science & technology **10**(6), 32–32 (2021)
19. Rivail, A., Schmidt-Erfurth, U., Vogl, W.D., Waldstein, S.M., Riedl, S., Grechenig, C., Wu, Z., Bogunovic, H.: Modeling disease progression in retinal octs with longitudinal self-supervised learning. In: International Workshop on PRedictive Intelligence In MEdicine. pp. 44–52. Springer (2019)
20. Russakoff, D.B., Lamin, A., Oakley, J.D., Dubis, A.M., Sivaprasad, S.: Deep learning for prediction of amd progression: a pilot study. Investigative ophthalmology & visual science **60**(2), 712–722 (2019)
21. Veeling, B.S., Linmans, J., Winkens, J., Cohen, T., Welling, M.: Rotation equivariant cnns for digital pathology. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II 11. pp. 210–218. Springer (2018)
22. Xiao, T., Wang, X., Efros, A.A., Darrell, T.: What should not be contrastive in contrastive learning. In: International Conference on Learning Representations (2020)
23. Xu, D., Xiao, J., Zhao, Z., Shao, J., Xie, D., Zhuang, Y.: Self-supervised spatiotemporal learning via video clip order prediction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10334–10343 (2019)
24. Yan, Q., Weeks, D.E., Xin, H., Swaroop, A., Chew, E.Y., Huang, H., Ding, Y., Chen, W.: Deep-learning-based prediction of late age-related macular degeneration progression. Nature machine intelligence **2**(2), 141–150 (2020)
25. Yim, J., Chopra, R., Spitz, T., Winkens, J., Obika, A., Kelly, C., Askham, H., Lukic, M., Huemer, J., Fasler, K., et al.: Predicting conversion to wet age-related macular degeneration using deep learning. Nature Medicine **26**(6), 892–899 (2020)
26. Zbontar, J., Jing, L., Misra, I., LeCun, Y., Deny, S.: Barlow twins: Self-supervised learning via redundancy reduction. In: International Conference on Machine Learning. pp. 12310–12320. PMLR (2021)
27. Zhao, Q., Liu, Z., Adeli, E., Pohl, K.M.: Longitudinal self-supervised learning. Medical Image Analysis **71**, 102051 (2021). https://doi.org/https://doi.org/10.1016/j.media.2021.102051, https://www.sciencedirect.com/science/article/pii/S1361841521000979