Self-supervised learning of phenotypic representations from cell images with weak labels

UNIVERSITY OF CAMBRIDGE

Jan Oscar Cross-Zamirski[1,2], Guy Williams[2], Elizabeth Mouchet[2], Carola-Bibiane Schönlieb[1], Riku Turkki[2], Yinhai Wang[2]

[1]DAMTP, University of Cambridge, [2]Discovery Sciences, R&D, AstraZeneca

AstraZeneca

Abstract

We propose WS-DINO as a novel framework to use weak label information in learning phenotypic representations from high-content fluorescent images of cells. Our model is based on a knowledge distillation approach with a vision transformer backbone (DINO), and we use this as a benchmark model for our study. Using WS-DINO, we fine-tuned with weak label information available in high-content microscopy screens (treatment and compound) and achieve state-of-the-art performance in not-same-compound mechanism of action (MOA) prediction on the BBBC021 dataset (98%), and not-same-compound-and-batch performance (96%) using the compound as the weak label. Our method bypasses single cell cropping as a pre-processing step, and using self-attention maps we show that the model learns structurally meaningful phenotypic profiles

Methods – DINO

DINO (Caron *et al.* 2021) is a self-supervised learning approach which incorporates aspects of knowledge distillation. Uniquely combining self-supervised learning with a Vision Transformer (ViT-S/8) backbone, DINO can learn features shown to perform well at k-NN clustering tasks.

The student network g_{θ_s} is trained with set of image crops to match the teacher network g_{θ_t}, which sees a different set of crops. For each image x a set of views V is generated which contains the two global crops $x^{g,1}$ and $x^{g,2}$ as well as eight local crops. The student is passed the set of global and local crops V, while the teacher sees only the global crops. The student network is trained to maximize the agreement between the outputs of g_{θ_s} and g_{θ_t}. To achieve this, the parameters of the student network θ_s are found by minimizing the cross-entropy loss:

$$\min_{\theta_s} \sum_{x \in \{x^{g,1}, x^{g,2}\}} \sum_{\substack{x' \in V \\ x' \neq x}} H(P_t(x), P_s(x'))$$

Where $H(a, b) = -a \log b$ and P_s and P_t are the probability distributions of the student and teacher respectively. Teacher weights θ_t are frozen during each epoch of student training and updated iteratively with an exponential moving average based on the previous weights of the student network with the formula: $\theta_t \leftarrow \lambda\theta_t + (1-\lambda)\theta_s$, where λ is the momentum parameter.

Both the student and the teacher network have a ViT backbone. The attention mechanism allows the ViT to synthesize information across the whole image as self-attention layers in the Transformer globally update the attention token embeddings.

Methods – WS-DINO

The **W**eakly **S**upervised form of self-**di**stillation with **no** labels (WS-DINO) is an adaptation to DINO.

We introduce the notation x_{i,y_i} to represent the i^{th} field of view of a fluorescent channel in the dataset which has been treated with treatment or compound y_i - the weak label. The superscript contains the crop information: g for global and l for local crops. When generating the sets of different views V_t (seen by teacher) and V_s (seen by student) for training, we enforce the constraint that the global and local crops are sampled from different images with the same weak label. We define the sets V_t and V_s for the randomly sampled ordered pair (i, j) where $i \neq j$ and $y_i = y_j$:

$$V_t = \{x_{i,y_i}^{g,1}, x_{i,y_i}^{g,2}\} \qquad V_s = V_t \cup \left\{x_{j,y_j}^{l,k} : \quad k \in \{1, ..., n\}\right\}$$

Where the superscript details the k^{th} local crop of n total crops (default: $n = 8$). Sampling the local crops from a different image is in contrast to sampling random crops with different augmentations from the same image, which is the method of DINO. For WS-DINO we minimize the loss:

$$\min_{\theta_s} \sum_{x \in V_t} \sum_{\substack{x' \in V_s \\ x' \neq x}} H(P_t(x), P_s(x'))$$

Experiments

We evaluated our experiments with the publicly available BBBC021 dataset, which consists of MCF-7 breast cancer cells exposed to several chemical compounds and stained with three fluorescent labels: DNA, F-actin, and β-tubulin. For training and evaluation, we used the subset which has been labelled for MOA evaluation (Ljosa *et al.* 2013) which consists of 12 unique MOA, 38 unique compounds and 103 unique treatments across 10 experimental batches.

We evaluated the corrected and aggregated feature embeddings (from the ViT backbone) with two metrics**: NSC matching** is a 1-Nearest-Neighbour match for each given treatment to the nearest neighbour in representation space which is not of the same compound. Cosine distance is used as the distance measure. **NSCB matching** is the same method as NSC matching with the additional constraint to exclude treatments from the same compound and batch as the given treatment.

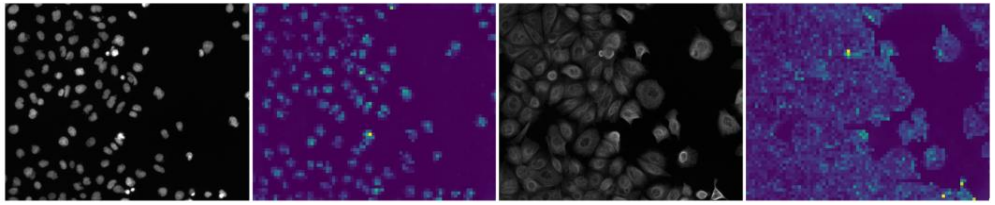

Figure 1: Self-attention maps from the ViT backbone of the network reveal the segmentation properties of the algorithm. Such visualizations increase confidence in the model by demonstrating the network is learning biologically and structurally meaningful features. Left: DNA, Right: F-Actin.

Results

Model	Weak Label	NSC	NSCB
DINO with ImageNet weights only	None	91%	82%
DINO finetuned on BBBC021	None	92%	95%
WS-DINO finetuned on BBBC021	Treatment	92%	90%
	Compound	**98%**	**96%**
	MOA	100%	100%

Table 1: WS-DINO achieves state-of-the-art results on BBBC021 using the compound as the weak label, **outperforming all known previous approaches using this dataset in NSC and NSCB MOA prediction** (including Janssens *et al.* 2020 (97/85%) & Perakis *et al.* 2021 (96/95%))**.**

We include MOA as a *pseudo-weak* label as a proof-of-concept for our method. One advantage of WS-DINO is that it is adaptable to datasets with partial MOA labels, a feature of some real drug discovery datasets.

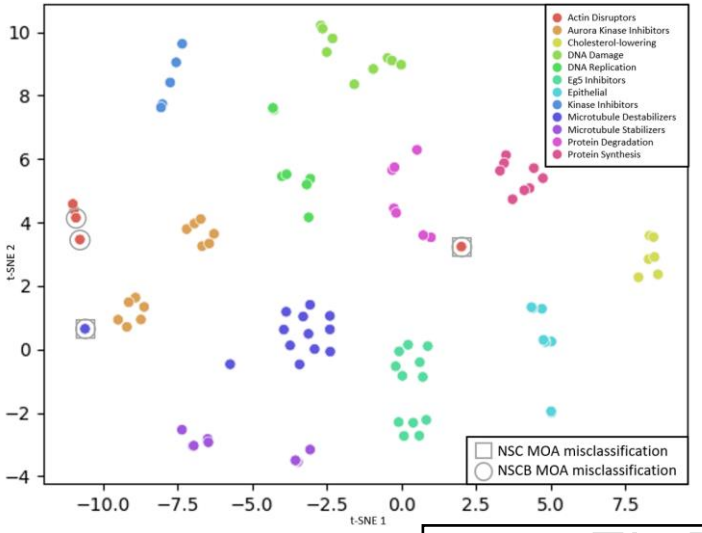

Figure 2: Two-dimensional t-SNE plot of each aggregated treatment feature from 200 epochs of training WS-DINO with compound as the weak label

For implementation details, the full paper, code, references and acknowledgements, scan the QR code!