Computational and Structural Biotechnology Journal 27 (2025) 928–936

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Research article

uniDINO: Assay-independent feature extraction for fluorescence microscopy images

Flavio M. Morelli [a,b,*], Vladislav Kim [a], Franziska Hecker [c], Sven Geibel [d], Paula A. Marín Zapata [a]

[a] R&D Machine Learning Research, Bayer AG, Pharmaceuticals Division, Berlin, Germany
[b] Department of Mathematics and Computer Science, Free University of Berlin, Berlin, Germany
[c] Proteome and Metabolome Research, Bielefeld University, Bielefeld, Germany
[d] R&D Hit Discovery, Bayer AG, Crop Science Division, Monheim, Germany

ARTICLE INFO

Keywords:
High-content imaging
Fluorescence microscopy
Morphological profiling
Self-supervised learning
Representation learning
Deep learning
Computer vision

ABSTRACT

High-content imaging (HCI) enables the characterization of cellular states through the extraction of quantitative features from fluorescence microscopy images. Despite the widespread availability of HCI data, the development of generalizable feature extraction models remains challenging due to the heterogeneity of microscopy images, as experiments often differ in channel count, cell type, and assay conditions. To address these challenges, we introduce uniDINO, a generalist feature extraction model capable of handling images with an arbitrary number of channels. We train uniDINO on a dataset of over 900,000 single-channel images from diverse experimental contexts and concatenate single-channel features to generate embeddings for multi-channel images. Our extensive validation across varied datasets demonstrates that uniDINO outperforms traditional computer vision methods and transfer learning from natural images, while also providing interpretability through channel attribution. uniDINO offers an out-of-the-box, computationally efficient solution for feature extraction in fluorescence microscopy, with the potential to significantly accelerate the analysis of HCI datasets.

1. Introduction

High-content imaging (HCI) combines advanced fluorescence microscopy with automated image analysis to extract quantitative data from cells and tissues, enabling the systematic investigation of disease mechanisms and the evaluation of therapeutic interventions at the cellular level. Effective feature extraction algorithms are crucial for HCI analysis, especially in the field of morphological profiling, where image features are used to characterize cellular responses to stimuli such as drug treatments and genetic modifications [1]. However, despite the vast availability of HCI data, there is a lack of general representation models which are applicable across datasets. This challenge likely arises from the heterogeneity of fluorescent images, which vary in channel count, fluorophore types, resolution, cell lines, and subcellular targets [2,3].

Numerous approaches have been proposed to extract feature representations from HCI experiments, primarily in the context of morphological profiling using the Cell Painting assay [4]. Traditional methods depend on customized cell segmentation and feature extraction pipelines, which are computationally demanding and require parameter adjustments for new data [1,5]. The first deep learning approaches used ImageNet-trained models to generate channel-wise embeddings that were subsequently concatenated [6]. However, transfer learning from natural images might not capture the domain-specific particularities of fluorescence microscopy data [7]. Other approaches have employed weakly supervised learning (WSL) to train CNN classifiers on noisy labels from experimental metadata [8,9], with adaptations proposed to reduce confounding from technical artifacts [10]. Nonetheless, these methods are sensitive to the quality of the noisy labels and require dataset curation to enhance performance.

Self-supervised learning (SSL) offers a promising framework for extracting feature representations in large-scale microscopy datasets while eliminating the need for extensive data curation. Following its success in natural images [11–16], SSL has gained traction in morphological profiling. The DINO framework [12] has been applied both at whole-image [17] and single-cell level [18], with several adaptations that leverage weak labels to mitigate confounding from experimental batches [19–21]. Alternative SSL paradigms which have proven

* Corresponding author at: R&D Machine Learning Research, Bayer AG, Pharmaceuticals Division, Berlin, Germany.
E-mail address: flavio.morelli@bayer.com (F.M. Morelli).

https://doi.org/10.1016/j.csbj.2025.02.020
Received 20 November 2024; Received in revised form 11 February 2025; Accepted 19 February 2025
Available online 24 February 2025

effective in fluorescent microscopy are variational autoencoders [22], masked autoencoders [3,17,23] and contrastive learning [17,24]. While these methods have achieved state-of-the-art results for Cell Painting images, most of them are not easily reusable for other assays with different microscopy settings and channel configurations.

Several approaches have been developed for fluorescence microscopy images with variable channel counts. Some of these methods concentrate in introducing modifications to the Vision Transformer (ViT) [25] backbone to generate embeddings for a varying number of channels. For example, Channel-ViT creates patch tokens independent from the channels while introducing a learnable channel embedding and dropping a subset of channels during training [26]. Similarly, ChAda-ViT employs inter-channel attention to embed biological images with an arbitrary number of channels [27]. Finally, DiChaViT extends the principle behind Channel-ViT and ChAda-ViT by introducing a sampling strategy to select the most diverse channels and adding channel and token diversification losses [28]. The embeddings produced by these methods combine information from all channels; however, single-channel embeddings are often desirable for interpretability in downstream biological tasks.

To address the need for single-channel disentanglement, several approaches have followed a concatenation strategy, extracting single-channel embeddings and subsequently concatenating them. An early example in this direction is CytoImageNet, which averages channels to create single-channel grayscale images for training, while extracting channel-wise embeddings during inference [29]. CA-MAE uses a single encoder with multiple channel-specific decoders during training and creates embeddings channel-wise with the trained encoder [23]; however, the reliance on channel-specific decoders limits its ability to train on images from highly diverse fluorescence microscopy assays. Similarly, Microsnoop adopts a single-channel masked autoencoder strategy that generates channel-wise embeddings followed by concatenation; however, it is trained on a relatively low number of cell images (ca. 10,000) and incorporates datasets beyond fluorescence microscopy [3].

Aiming to develop a more generalizable model that effectively accommodates the diversity of HCI data, we introduce uniDINO, a feature extraction model specifically designed for fluorescence microscopy images with an arbitrary number of channels. uniDINO is trained on a collection of over 900,000 single-channel images from various assays, using the DINO SSL framework and employing single-channel feature concatenation to generate multi-channel embeddings. We evaluate our model across multiple biological applications, assessing its generalizability on unseen data and its robustness against technical artifacts commonly found in HCI. Our key contributions are:

1. We provide a general-purpose model for feature extraction in microscopy images trained on a large corpus of HCI assays.
2. We benchmark our model extensively on diverse datasets encompassing different species, cell lines and markers, demonstrating state-of-the-art performance and generalizability across assays.
3. We illustrate how single-channel embeddings can be used for biological interpretation.

We provide model weights under a non-commercial license, anticipating that they will be used by the scientific community to accelerate research in high-content imaging (HCI) data and cellular profiling specifically.

2. Materials & methods

2.1. Data

We train and evaluate uniDINO on diverse datasets generated by fluorescence microscopy assays with a varying number of channels. The assays encompass various cell lines and biological tasks, such as mechanism of action (MoA) prediction, clustering of genetic perturbations, or assessing subcellular protein localization. Five datasets were used for training and three held-out datasets for evaluation (Table 1). Details on the data are included in Supplement A.

Training data includes two Cell Painting datasets (JUMP-CP screening plates [30] and BBBC037 [31]), one internal dataset (Cardio), as well as the BBBC021 [32] and Human Protein Atlas (HPA) [33] datasets, which are prominent in the HCI field. Our JUMP-CP set contains a subset of 3736 chemical perturbations from the publicly available JUMP-CP data screened in U2OS cells, as detailed in Supplement A.3. BBBC037 comprises gene overexpression perturbations of 190 genes, also screened in U2OS cells. The Cardio set examines iPSC-derived cardiomyocytes subjected to CRISPR knockout perturbations. This assay includes a sarcomere channel which displays a striated pattern distinctive of heart muscle cells. The BBBC021 dataset explores the impact of 111 compounds at 8 concentrations in MCF-7 cells, and reports MoA annotations for a subset of treatments. Finally, the HPA dataset features 17 cell lines and organelle labels for protein localization.

The evaluation set includes three Cell Painting datasets: JUMP-CP Target-2 plates, Cell Health and Insect. The JUMP-CP Target-2 sentinel plates were also generated by the JUMP-CP consortium [30], but were not included in the training set. The Insect dataset [34] includes 36 chemical perturbations across seven concentrations in the Sf9 insect cell line, thus allowing exploration of model performance on non-human cells. The Cell Health data assesses the predictive capabilities of Cell Painting features on cell health assays measuring apoptosis, proliferation, DNA damage, and cell cycle stage [35]. Together, these evaluation sets aim to investigate the robustness and applicability of the model to unseen data across different biological contexts.

2.2. uniDINO training and inference

Our method employs DINO [12] to extract features from single-channel images during training, followed by the concatenation of single channel embeddings at inference time (Fig. 1). Previous work has shown DINO to have superior performance for HCI [17]. By training on individual channels, we can leverage a diverse array of datasets while eliminating the need for separate projection layers for images with different channel configurations. After pretraining uniDINO on five datasets (Table 1), we use the encoder (small vision transformer or ViT-S) as a feature extractor.

2.2.1. Model architecture and training

We train a single-channel ViT-S on image crops without segmentation using the self-supervised DINO framework (Fig. 1). During training, a random microscopy channel is selected and image crops of size 224x224 pixels are provided as input to the encoder. We train the model for 100 epochs with a batch size of 510. We use the AdamW optimizer with a learning rate of $4 \cdot 10^{-4}$ and a cosine schedule to decrease it to $1 \cdot 10^{-6}$ at the end of training. An additional cosine schedule is used to increase the weight decay from 0.04 to 0.4. We warm up the learning rate linearly for 20 epochs and DINO's teacher temperature for 30 epochs from 0.01 to 0.04. The momentum for the DINO teacher is 0.9995. The DINO projection head dimensionality is 20,000. All hyperparameters are chosen based on a related study comparing SSL methods for Cell Painting [17]. Adjustments in batch size, warmup of the learning rate and the teacher temperature are applied to prevent divergence during training. During training we use the 'flip' and 'color' augmentations described in [17]. Given the different lengths of the datasets, we cycle the smaller ones to match the length of the largest dataset, so that all datasets contribute equally during training. This cycling approach means that the model is exposed to approximately 4.1 million images per epoch. The model is trained using five NVIDIA Tesla V100 GPUs, each with a VRAM of 32 GB, over a period of approximately two weeks.

Table 1
List of training and evaluation datasets. Overview of fluorescence microscopy datasets used in this study, including channels, cell lines, and available annotations. Dataset sizes report the number of single-channel images and are rounded to the nearest multiple of 1000.

Dataset	Channels	Annotations	Cell Line	Size	Usage
JUMP-CP screening plates	DNA, RNA, ER, AGP, Mito	Compound / Gene target	U2OS	**850k**	Training
BBBC037	DNA, RNA, ER, AGP, Mito	Gene perturbation	U2OS	**23k**	Training
Cardio	DNA, Sarcomere, other channels	Gene perturbation	iPSC-derived cardiomyocytes	**8k**	Training
BBBC021	DNA, Actin, Tubulin	Compound / MoA	MCF−7	**5k**	Training
HPA	DNA, ER, Tubulin, Protein	Cell line / Organelle	17 human cell lines	**18k**	Training
JUMP-CP Target−2 plates	DNA, RNA, ER, AGP, Mito	Compound / Gene target	U2OS	task dependent	Evaluation
Insect	DNA, RNA, ER, AGP, Mito	Compound / MoA	Sf9 (insect)	task dependent	Evaluation
Cell Health	DNA, RNA, ER, AGP, Mito	Cell health readouts	U2OS	task dependent	Evaluation

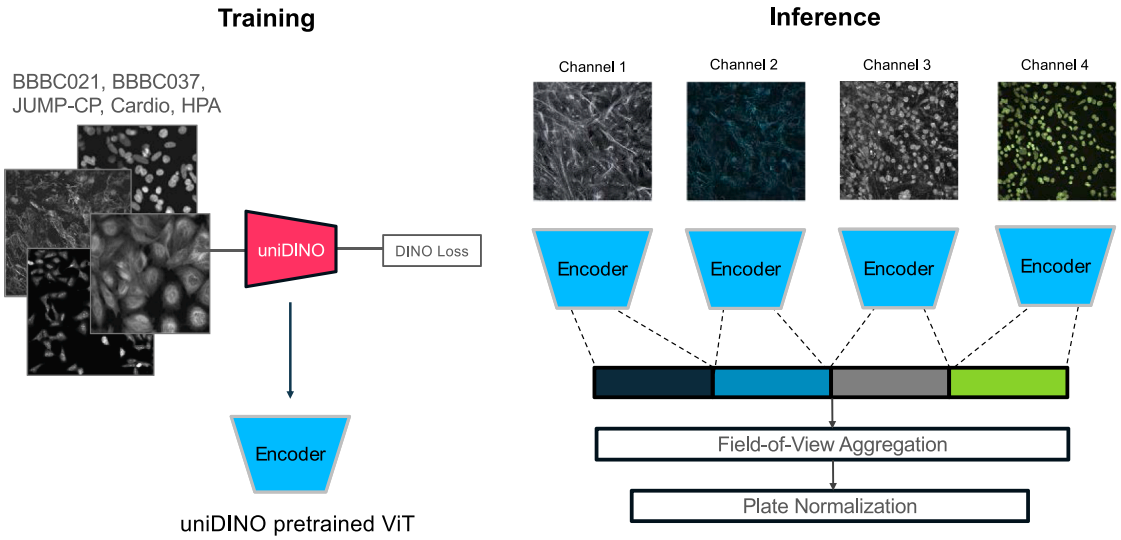

Fig. 1. uniDINO training and inference. a) During training, a collection of single channels picked randomly from multi-channel images is loaded as a training minibatch and passed to DINO. Smaller datasets are repeated to match the size of the longest dataset in each epoch. b) For inference in a multi-channel image, embeddings of each single channel are obtained with the pretrained ViT teacher network and subsequently concatenated to create an image embedding. Whenever multiple fields-of-view (FOV) per well are available, the image embeddings are averaged across FOVs. The aggregated well embeddings are plate-normalized (see Section 2.2.2.).

2.2.2. *Inference and postprocessing*

To generate embeddings, we use the pretrained teacher ViT from DINO. During inference, each channel of a single image is divided into crops of 224x224 pixels. Embeddings for each image crop are computed and averaged into the channel-specific mean. This approach of dividing each image into smaller crops offers computational efficiency without compromising performance [17]. Finally, channel-specific mean embeddings are concatenated to form the whole-image embedding. To generate well profiles, we average across all fields-of-view (FOV) images in a well.

A crucial step in cellular profiling is embedding postprocessing, as normalization methods can have a significant impact on the performance of image features in downstream tasks. Following previous work [17], we adopt plate normalization as our postprocessing strategy. Specifically, we use MAD robustize for a given plate defined as

$$h_{norm} = \frac{h_{well} - med(h_{well})}{MAD(h_{well})},$$

where h_{well} is the well profile, *med* is the plate median and *MAD* is the median absolute deviation. In absence of plate information, we do not perform any postprocessing.

2.3. *Baselines and benchmarking protocol*

2.3.1. *Baseline methods selected for comparison*

We compare uniDINO embeddings with 5 baseline feature extractors: Microsnoop, a channel-agnostic feature extractor for microscopy images [3]; a model pretrained on natural images (ImageNet DINO); CellProfiler [36]; a ViT-S with randomly initialized weights (Random ViT); a random Gaussian baseline. To make the results comparable with uniDINO, we use Microsnoop only in 'tile mode', which can lead to inferior results on certain datasets with respect to using the 'cell region cropping mode' [3]. For transfer learning, we use DINO ViT-S trained on ImageNet-1k (https://huggingface.co/timm/vit_small_patch16_224.dino) to generate features in a channel-wise basis by replicating each channel to fit the RGB format. This serves as a baseline to determine the benefits of training exclusively on fluorescence microscopy images. We use the pretrained weights without finetuning. The randomly initialized ViT model allows to check the extent to which a random projection reveals the intrinsic structure of the data. For CellProfiler, we perform feature selection after plate normalization, while for the other models, we directly use plate-normalized embeddings.

2.3.2. *Evaluation metrics for classification tasks*

To assess the performance of the embeddings on downstream biological tasks, we use a nearest-neighbor classifier with restrictions on the possible match. For this, we use annotations in each dataset, such as perturbation or MoA, as target variables. Given the unique characteristics of each assay, we adapt the evaluation metrics to better suit both the structure and available annotations of each dataset.

For BBBC021, we assess the performance of our embeddings in classifying mechanism of action (MoA) and compound annotations. We focus exclusively on embeddings with MoA annotations and include

only those MoAs that involve at least two different compounds. MoA classification is evaluated using the not-same-compound (NSC) accuracy and not-same-compound-and-batch (NSCB) accuracy, two commonly used metrics for this dataset [8,37]. NSC accuracy is calculated based on both mean treatment embeddings and well embeddings, where a treatment is defined as a unique combination of compound and concentration. This metric restricts nearest-neighbor matches to data points from different compounds. Since embedding similarities might be higher between samples screened within the same batch, we use the NSCB accuracy to exclude intra-batch matches. This metric is calculated based on well embeddings and further restricts matches to data points from different compounds and batches. A smaller gap between NSCB and NSC accuracies indicates a better ability to capture biological signal in the presence of technical artifacts, specifically batch effects. For NSCB accuracy, we remove MoAs that are only present in one batch. Finally, we assess the performance on compound classification using the not-same-plate-and-well (NSPW) accuracy. This metric is computed based on well profiles and restricts matches to different plates and well locations, thereby reflecting potential well positioning effects.

The held-out JUMP-CP Target-2 dataset consists of 27 plates from 16 batches screened in four different laboratories (denoted as sources). A subset of active compounds is selected, as detailed in Supplement A.3. Our analysis involves predicting both the gene target annotation to assess the biological signal present in the profiles and the compound label to evaluate reproducibility. First, we use batch-aggregated compound profiles to determine both NSC and NSCB accuracies when predicting gene target. The drop between NSC and NSCB accuracies gives a sense of the impact of batch effects on the performance in downstream tasks. Lastly, we predict compound perturbation using well profiles and report both not-same-batch (NSB) and not-same-source (NSS) accuracies to assess batch and source effects. Each plate contains only one replicate of each compound, and all 27 share an identical layout, thus making it impossible to check for positional effects.

Finally, for the Insect dataset, which consists of six plates from a single batch, we evaluate classification performance on MoA as well as compounds annotations. We use both profiles aggregated at the treatment level (compound/concentration) and well profiles to calculate metrics. We use NSC accuracy after predicting the MoA annotation. For MoA prediction, we only consider MoAs that include at least two different compounds, and we merge the annotations of mitochondrial complex inhibitors I, II and III into a single MoA. The metrics are calculated using a subset of active concentrations, with further details provided in Supplement A.3. Moreover, we introduce the not-same-compound-or-plate (NSCP) accuracy, which excludes wells either from the same plate or treated with the same compound. A lower drop between NSC and NSCP accuracy indicates better robustness to technical artifacts. Nevertheless, relying only on MoA classification can be misleading, as the distribution of compounds across MoAs is highly unequal. Therefore, we also calculate the accuracy in the classification of compound labels in a not-same-plate-or-well (NSPW) fashion, which reflects plate and well positional effects.

2.3.3. Evaluation metrics for cell health predictions

In addition to classification, we further evaluate our embeddings on multiple regression tasks using the Cell Health dataset, where Cell Painting features have been leveraged to predict cell health assays [35]. We adapt the analysis pipeline from [35] by using random forests instead of ElasticNet. Random Forest is robust and versatile in handling different datasets and is not limited by a linear functional form, as is the case with ElasticNet. We employ the random forest implementation in scikit-learn, version 1.0.2, [38], which consists of 100 trees, and assess the predictive power of the trained model with 5-fold cross-validation.

To condense metrics across tasks, we define a new metric based on the R-squared performance of predicting each cell health readout. We calculate the R-squared on the held-out dataset and clip all the negative values to zero, averaging across the five folds. We then calculate the proportion of tasks that have an R-squared equal or higher than a threshold t. We calculate these proportions for multiple thresholds in the range (0, 1), which gives a curve similar to a Kaplan-Meier curve:

$$S_t = \frac{1}{K}\sum_{k=1}^{K} I\left[R_k^2 \geq t\right], \qquad t \in \{t_1, \ \ldots, \ t_M\}, \quad 0 \leq t_1 < \ldots < t_M \leq 1 \ ,$$

where t is a threshold between 0 and 1, K is the number of cell health tasks (70 in total), R_k^2 is the R-squared for task k, M is the number of thresholds for which the score is calculated and I is the indicator function. We take the area under the curve (AUC) as a measure for model predictive power, with a higher AUC indicating better performance. The AUC of the score is defined as

$$AUC_S = \int_0^1 S_t dt,$$

which is approximated using the trapezoidal rule and the M discretized thresholds. As S_t is between 0 and 1, AUC_s will also be between 0 and 1.

2.4. Explainability through single-channel feature importance

A key advantage of uniDINO is the ability to assign extracted features to a specific channel, which provides interpretability in downstream biological tasks. To showcase this for MoA inference, we train binary random forest classifiers to predict a selected class against all other classes within a given dataset and use the feature importance scores to analyze the distribution of the top 50 features across channels. Random forests are trained using 100 components and classes are balanced with their inverse frequency through class weighting.

3. Results

We first evaluate uniDINO embeddings using the BBBC021 data from the training set. Evaluation is based on the classification of biological annotations, using a different task from that which the model was originally optimized for. For BBBC021, we additionally showcase how uniDINO features can yield interpretable results by leveraging the association of each feature with a specific channel. Following this, we test uniDINO on unseen evaluation datasets. We use the JUMP-CP Target-2 plates to check performance on unseen data generated under similar conditions to the training set, the Insect dataset to assess generalization in non-mammalian cells, and the Cell Health dataset to evaluate the predictive capabilities of our embeddings in cell health assays. For all datasets, we compare uniDINO embeddings against various feature extractors, including Cell Profiler, transfer learning from ImageNet (ImageNet DINO), Microsnoop, and two random baselines, as detailed in Section 2.3.1.

3.1. uniDINO features enable mechanism of action identification in BBBC021

First, we use uniDINO to extract image features from the BBBC021 dataset, which consists of compound perturbations imaged across three fluorescent channels (Fig. 2a). This dataset includes manually curated MoA annotations, providing reliable labels for performance estimation and feature interpretation. The embeddings are evaluated based on the classification of MoA and compound labels, using multiple metrics to also investigate batch effects (see Section 2.3.2.).

Classification accuracies show that uniDINO outperforms other approaches in all metrics (Table 2). Although all models experience a decrease in accuracy when enforcing cross-batch MoA matching (NSC vs. NSCB), uniDINO exhibits the smallest drop (-12.4 %), followed by CellProfiler (-16.8 %), ImageNet DINO (-30.2 %) and Microsnoop

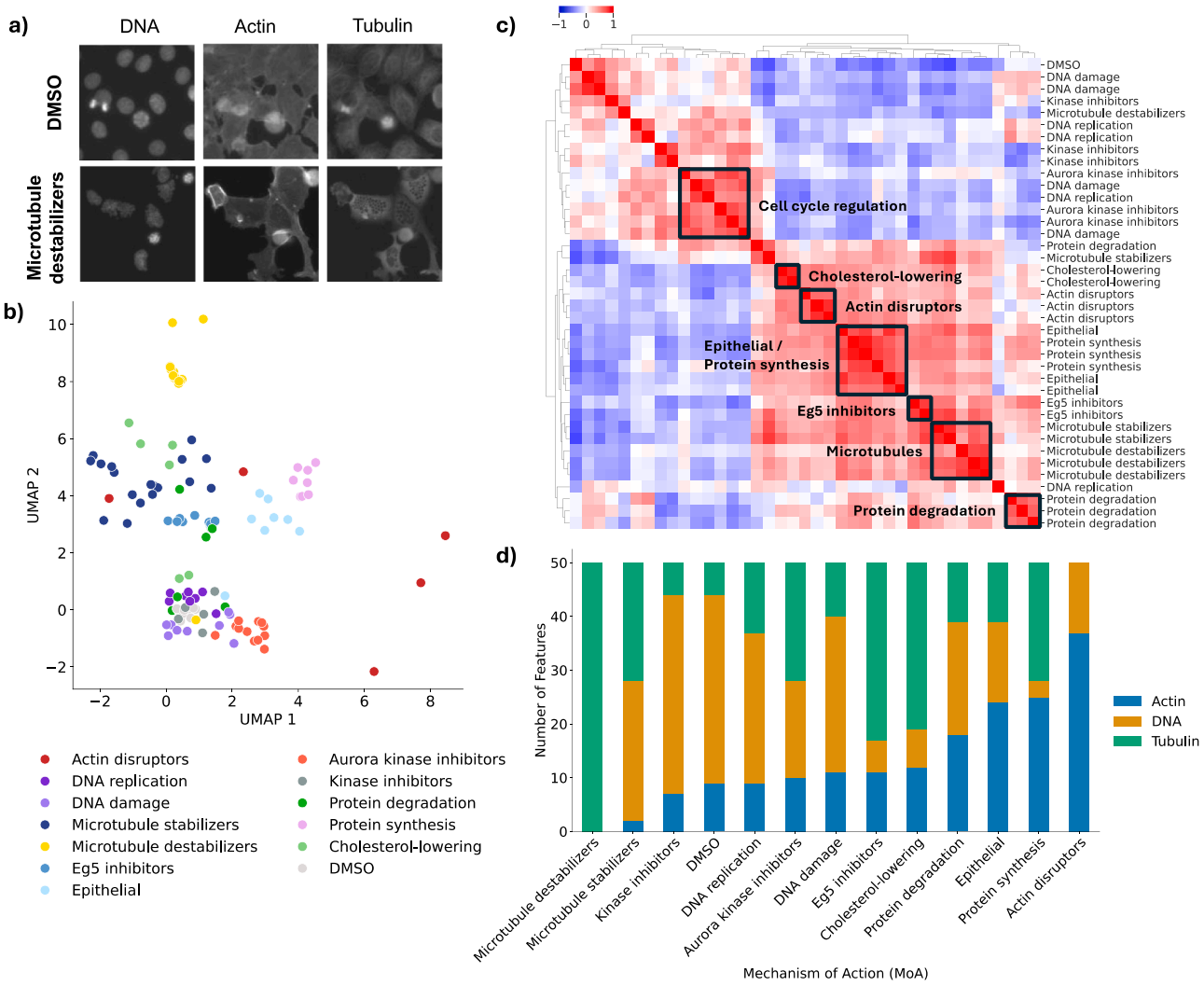

Fig. 2. Exploration of uniDINO features from the BBBC021 dataset. a) Representative images from selected MoAs. b) UMAP projections of features aggregated per treatment (compound and concentration) for treatments with MoA annotations. Colors highlight different MoAs. c) Hierarchical clustering of profiles aggregated per compound. d) Distribution of the top 50 most important features across fluorescent channels to classify each MoA against all other classes (see Section 2.4.).

Table 2
Nearest-neighbor classification of MoA and compound annotations in BBBC021 dataset. Not-same-compound (NSC), not-same-compound-or-batch (NSCB) and not-same-plate-or-well (NSPW) accuracy are reported. The parentheses indicate the predicted label (MoA or compound) and the aggregation level of the profiles (treatment or well). N indicates the number of profiles used for classification. The best results are highlighted in bold.

	NSC accuracy (MoA/ treatment) N = 103	**NSC accuracy (MoA/well)** N = 302	**NSCB accuracy (MoA/well)** N = 274	**NSPW accuracy (compound/ well)** N = 302
uniDINO	**0.845**	**0.795**	**0.697**	**0.656**
Microsnoop	0.485	0.477	0.226	0.457
ImageNet DINO	0.786	0.685	0.478	0.636
CellProfiler	0.573	0.596	0.496	0.646
Random ViT	0.417	0.381	0.157	0.450
Gaussian	0.039	0.050	0.047	0.017

(-52.6 %), and Random ViT (-58.8 %). Moreover, uniDINO shows the best performance on classifying compounds across plates and wells (NSPW). These results highlight the superiority of uniDINO features in both extraction of biological signal and robustness to technical artifacts.

Surprisingly, the accuracies of Random ViT are closer to those of Microsnoop (the worst-performing trained model) than to Gaussian noise, indicating that transformations applied by a randomly initialized ViT model still capture complex relationships in the data. However, these random representations show high susceptibility to technical artifacts, as Random ViT exhibited the greatest performance decline when subjected to cross-batch matching. The accuracies presented in Table 2 are lower than published results [3,6,37]. This discrepancy arises from our choice to minimize postprocessing of embeddings, using only plate normalization for a consistent comparison of models. Additional post-processing steps such as Typical Variation Normalization (TVN) [6] improve the performance of all feature extractors, while still maintaining uniDINO as the best-performing model (Supplement B).

We further explore uniDINO embeddings in the BBBC021 dataset qualitatively through 2D projections and clustering. UMAP projections [39] of treatment-level embeddings show separation of several MoAs, including ‘Microtubule destabilizers’, ‘Protein synthesis’, ‘Aurora kinase inhibitors’, ‘Epithelial’, and ‘Cholesterol lowering’, while the remaining MoAs are indistinguishable from DMSO or are highly overlapping (Fig. 2b). Additionally, we aggregate embeddings by compound across active concentrations and create a hierarchical clustering similarity map (Fig. 2c), which is a standard approach in the field [40,41]. This map confirms the clustering of several MoAs observed in Fig. 2b

(microtubule-related, 'Protein synthesis', 'Eg5 inhibitors') and reveals further groupings which were not clearly resolved by UMAP plots ('Actin disruptors', 'Protein degradation', 'Cholesterol-lowering'). A large group of compounds clustering with DMSO is also observed (top-left), suggesting that these compounds are phenotypically inactive. The UMAP and similarity map results from uniDINO are comparable with the results from ImageNet DINO but exhibit a clearer separation of MoAs than Microsnoop and CellProfiler (see Supplement C).

Finally, we leverage uniDINO single-channel embeddings to assess the relative importance of individual channels in the classification of MoAs. To achieve this, we use feature importance scores from random forest classifiers as detailed in Section 2.4. The distribution of top features among channels shows expected associations between the most relevant channels and their respective MoAs (Fig. 2d). For example, tubulin is the most frequent channel for 'Microtubule destabilizers', as is the actin channel for 'Actin disruptors' and the DNA channel for 'DNA damage' and 'DNA replication'.

In summary, the results from the BBBC021 dataset show that uniDINO effectively captures biological signals relevant for MoA identification. Additionally, it exhibits robustness against batch effects and enables the quantification of channel importance in downstream tasks.

3.2. uniDINO performance in the JUMP-CP dataset

We continue our analysis by extracting embeddings for the sentinel (Target-2) plates from the JUMP-CP dataset, which comprises compound perturbations imaged with the Cell Painting assay across different laboratories (sources). Although this subset of plates was not part of the training set, they share similarities with the training data, as they were imaged in the same laboratories and batches. We use classification metrics for both the compound and gene target labels for a subset of active compounds (see Section 2.3.2).

CellProfiler features outperform all other methods in classifying gene targets based on NSC and NSCB accuracies (Table 3), with DINO ranking as the second best. However, given the low absolute accuracies, the results for target prediction should be interpreted with caution. Possible reasons for these low values include unreliability of target annotations, which were not as thoroughly curated as those in the BBBC021 dataset, and the limitation of matching each compound to only one target, which does not consider potential poly-pharmacology. Additionally, over 70 % of compounds are phenotypically indistinguishable from controls ('inactive'), which leads to low prediction accuracies.

Additionally, CellProfiler displays the highest accuracy in classifying compound annotations across batches and laboratory sources (NSB and NSS), contrasting with the superior performance of uniDINO in cross-batch classification tasks in BBBC021 data. The drop in performance relative to CellProfiler is particularly pronounced for the cross-source matching task across all models. This result may arise from the fact that, for the JUMP-CP dataset, the CellProfiler pipeline was specifically tailored to optimize feature extraction for each individual source, a process that was not applied to the other models. The substantial drop between NSB as NSS accuracies indicates pronounced differences between samples screened at various laboratories. This discrepancy highlights the challenges of integrating embeddings across different sources and emphasizes the need for further postprocessing, as described in [42]. Notably, unlike other datasets that measure each compound in multiple well locations, the same compound is consistently screened in the same well in the Target-2 plates. Therefore, it is not possible to assess whether the compound classification performance was influenced by plate-positional effects.

In summary, CellProfiler delivers the most informative features in the Target-2 plates from the JUMP-CP dataset. uniDINO achieves slightly lower performance but benefits from not requiring manual adaptation to process the different laboratory sources. However, the low target accuracies and the inability to assess well positioning effects could limit the reliability of the conclusions drawn from this data.

Table 3
Nearest-neighbor classification of gene target and compound annotations in Target-2 plates from the JUMP-CP dataset. Not-same-compound (NSC), not-same-compound-or-batch (NSCB), not-same-batch (NSB), and not-same-source (NSS) accuracy are reported. The parentheses indicate the predicted label (gene target or compound) and the aggregation level of the profiles (batch or well). N indicates the number of profiles used for classification. The best results are highlighted in bold.

	NSC accuracy (target/batch) N = 2599	NSCB accuracy (target/batch) N = 2599	NSB accuracy (compound/well) N = 4216	NSS accuracy (compound/well) N = 4216
uniDINO	0.029	0.017	0.431	0.045
Microsnoop	0.019	0.006	0.152	0.005
ImageNet DINO	0.028	0.017	0.418	0.075
CellProfiler	**0.032**	**0.023**	**0.491**	**0.111**
Random ViT	0.016	0.006	0.095	0.007
Gaussian	0.004	0.004	0.005	0.004

3.3. uniDINO generalizes to an unseen insect cell dataset

To further evaluate the generalizability of uniDINO to unseen data, we use the held-out Insect dataset (Fig. 3a). We report results for the classification of MoA and compound labels using several metrics that enable the analysis of plate and well positioning effects (see Section 2.3.2). The nearest-neighbor accuracy metrics from Table 4 indicate that uniDINO outperforms all other methods except for NSC accuracy at the treatment level, where ImageNet DINO has the highest value. uniDINO also shows the lowest accuracy drop when enforcing cross-plate MoA matching (-7.4 % NSC vs NSCP), as well as the highest compound classification accuracy matching across plates and wells (NSPW). Overall, these results highlight uniDINO's effectiveness in extracting biological information and the enhanced technical reproducibility of uniDINO embeddings compared to its counterparts.

Qualitative examination of uniDINO embeddings using UMAP reveals that most inhibitors primarily cluster according to their MoA (Fig. 3b). We observe a distinctive cluster of actin inhibitors and a supercluster of respiratory MoAs which groups together inhibitors of Complexes I, II, and III, mitochondrial ATPase as well as those of their close structural relative, the vacuolar ATPase, in line with previous reports [34]. However, other MoAs, such as tubulin inhibitors, form distinct clusters for different compounds, likely because these compounds (paclitaxel and colchicine) interact with their target in different ways [43]. Since UMAP plots include all tested concentrations, a partial overlap of some MoAs with negative controls might indicate inactive concentrations. Furthermore, the complete overlap of the neurotoxin MoA with negative controls reinforces the absence of neuronal signaling in Sf9 cells. The hierarchical clustering in Fig. 3c further supports MoA-based grouping, including distinct clusters for actin inhibitors, inactive compounds, and a larger group comprising respiration and cellular homeostasis. In contrast, other distinct MoAs, such as apoptosis inducers, show minimal similarity with the larger clusters of actin inhibitors and cellular respiration as expected.

Together, the evaluation of uniDINO embeddings on the Insect dataset confirms their ability to capture MoA information and demonstrates robust cross-species generalization.

3.4. uniDINO accurately predicts cell health readouts

Finally, we evaluate the predictive power of uniDINO features by assessing their performance in predicting cell health assay readouts, as previously demonstrated with CellProfiler features [35]. For this, we use the held-out Cell Health dataset consisting of Cell Painting images and cell health assays tested across three cell lines (A549, ES2, HCC44).

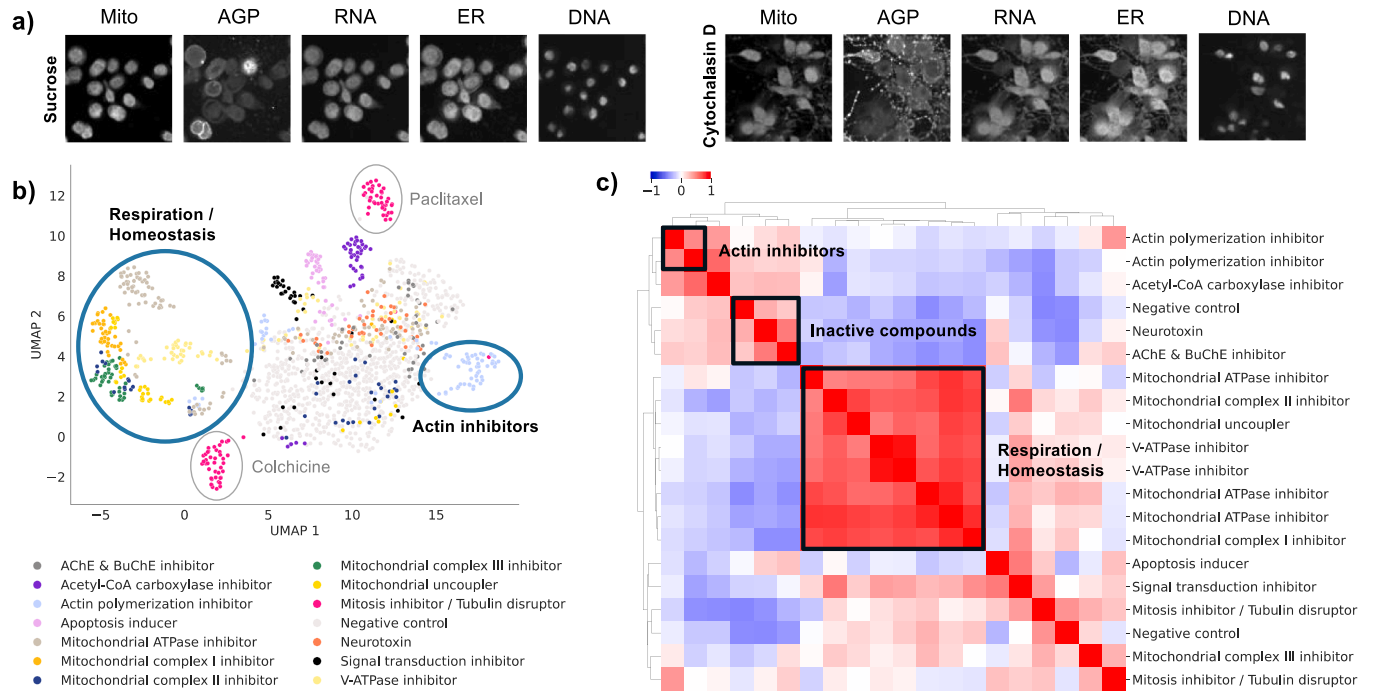

Fig. 3. Exploration of uniDINO features from the Insect dataset. a) Representative images from the Insect dataset for sucrose (negative control) and the actin inhibitor cytochalasin D, which exhibits protrusions from the cytoskeleton. b) UMAP projections of well-level features including all tested concentrations per compound. Dot colors highlight different MoAs. Blue circles indicate larger MoA groups, grey circles single compounds. c) Hierarchical clustering of compounds with MoA annotations for active concentrations based on profiles aggregated at the compound level.

Table 4
Nearest-neighbor classification metrics for the Insect dataset. We report not-same-compound (NSC), not-same-compound-or-plate (NSCP) and not-same-plate-or-well (NSPW) accuracy from nearest-neighbor classification. The parentheses indicate the predicted label (MoA or compound) and the aggregation level of the profiles (treatment or well). Moreover, we include the number of aggregated profiles used for classification. The best results are highlighted in bold.

	NSC accuracy (MoA/ treatment) N = 43	NSC accuracy (MoA/well) N = 258	NSCP accuracy (MoA/well) N = 258	NSPW accuracy (compound/ well) N = 258
uniDINO	0.605	**0.624**	**0.578**	**0.748**
Microsnoop	0.512	0.376	0.306	0.636
ImageNet DINO	**0.744**	0.605	0.543	0.632
CellProfiler	0.140	0.391	0.360	0.236
Random ViT	0.395	0.295	0.236	0.442
Gaussian	0.116	0.128	0.105	0.112

Regression metrics for all assays are aggregated into a single AUC score, as explained in Section 2.3.3.

The AUC metrics presented in Table 5 demonstrate that uniDINO achieves the best performance in A549 and HCC44 cell lines, while CellProfiler excels in ES2 cells. However, the performance differences are quite modest, with ImageNet DINO showing comparable AUC values. These results indicate that uniDINO performs on par with other models when predicting orthogonal assays on unseen data, further validating its utility as an out-of-the-box tool for high-content analysis.

Table 5
Predictive performance on cell health assays. AUC_s metric for cell health readouts on three different cell lines.

	A549	ES2	HCC44
uniDINO	**0.198**	0.260	**0.227**
Microsnoop	0.191	0.173	0.188
ImageNet DINO	0.194	0.230	0.209
CellProfiler	0.192	**0.270**	0.199
Random ViT	0.151	0.172	0.171
Gaussian	0.001	0.001	0.000

4. Discussion

This work presented uniDINO, a generalist model for feature extraction in fluorescence microscopy, which can handle images with any number of channels. Our single-channel feature concatenation strategy enabled us to train and evaluate our model across large, heterogeneous datasets, demonstrating that uniDINO outperforms baseline methods in predicting biological labels, generalizes across biological domains, and exhibits robustness against technical artifacts. Overall, uniDINO represents a significant advancement that accelerates cellular phenomics analysis across various organisms and experimental designs.

We trained uniDINO on a large, diverse set of approximately 900,000 single-channel fluorescent microscopy images, showing that training directly on HCI data provides superior results compared to transfer learning from natural images. Given its strong performance on held-out data, uniDINO can serve as an out-of-the-box feature extractor in fluorescence microscopy, eliminating the need for both finetuning to new datasets and training specialized models for different assays. This makes our model especially valuable in low-data scenarios. Additionally, it provides a convenient and computationally efficient alternative to software solutions that require manual tuning, such as CellProfiler. Although training uniDINO requires considerable resources (five GPUs over two weeks), the speed during inference is substantial. The computational efficiency is comparable to a related study focused on Cell Painting images [17], which reported a 50x reduction in average processing time and cost using a multi-channel DINO model compared to CellProfiler.

Our benchmark analysis indicates that uniDINO features have enhanced robustness to technical artifacts. For most datasets, it achieves

higher nearest-neighbors accuracy than Microsnoop, ImageNet DINO, and CellProfiler when classifiers are restricted to match only across different compounds, batches, plates, or wells. This improvement may stem from the diversity of the training data, which encompasses images from various laboratories, microscopes, staining techniques, organisms, and cell lines.

Related methods, including CA-MAE and Microsnoop, have addressed the same question of providing channel-agnostic feature extractors. However, CA-MAE was exclusively trained on Cell Painting data and relies on channel-specific decoders that limits its application to highly diverse datasets. Moreover, we observe that Microsnoop underperforms uniDINO in all evaluation tasks. This lower performance may be attributed to its reliance on segmentation for certain datasets, a small training set size, a low masking ratio, or the inclusion of non-fluorescent microscopy images in its training set.

Other model architectures using a channel-agnostic ViT backbone, such as Channel-ViT and ChAda-ViT, could also be used to train generalist models for high-content analysis. However, they produce image-level embeddings that provide a combined representation for all channels. We deliberately chose a single-channel embedding strategy to evaluate the importance of individual channels, as demonstrated with the BBBC021 dataset (see Section 3.1.). This approach offers valuable insights for biological discovery. A drawback of channel concatenation is the linear increase in embedding dimensionality with the number of channels. This issue could be alleviated by employing feature selection or dimensionality reduction algorithms on the multichannel or single-channel embeddings. Another key limitation of single-channel feature extraction is its failure to account for spatial correlations between channels, which hampers the model's ability to detect biologically meaningful colocalizations across different cellular compartments. In contrast, multi-channel feature extractors like [17,27] include such channel interactions by design. Notably, the multi-channel DINO model trained in [17] demonstrates superior performance for the JUMP-CP Target-2 plates compared to a single-channel DINO trained on a similar dataset (Supplement D), which highlights the trade-off between flexibility and performance.

While our work focused on applying uniDINO across heterogeneous data, our training and evaluation sets were predominantly based on Cell Painting images. Therefore, the model's performance could benefit from further diversifying the dataset; for example, by including images of yeast, neurons or filamentous fungi, which exhibit distinct morphologies. In Supplement E, we show that uniDINO exhibits signs of performance saturation with respect to Cell Painting data, as training solely on the JUMP-CP subset already yields top performance on Cell Painting evaluation sets. By contrast, performance on other datasets, such as BBBC021, benefits from the inclusion of the more diverse full training set.

Future research should explore the impact of scaling the ViT backbone [44] alongside increasing the size and diversity of the training data. Additionally, investigating whether low-cost finetuning strategies can enhance model performance on specific datasets remains an open question. Given that the principle of using a single-channel backbone in uniDINO can be integrated with other self-supervised learning (SSL) approaches, another promising area for future research is to benchmark alternative SSL methods, such as DINOv2, iBOT, and MAE, as demonstrated in previous studies on SSL for HCI [17]. Finally, our model is specifically designed for fluorescent images generated from cell cultures, where cells grow differently than in tissues. Future research should explore the applicability of uniDINO to related modalities such as immunofluorescence staining of tissue sections across multiple datasets and diseases, similar to other work in histopathology [45,46].

CRediT authorship contribution statement

Geibel Sven: Writing – review & editing. **Hecker Franziska:** Writing – review & editing, Resources, Data curation. **Kim Vladislav:** Writing – review & editing, Supervision, Software, Methodology. **Morelli Flavio:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Marín Zapata Paula Andrea:** Writing – review & editing, Supervision, Project administration, Methodology, Conceptualization.

Conflict of Interest

Flavio M. Morelli, Vladislav Kim, Sven Geibel, and Paula A. Marín Zapata are employees of Bayer AG. Franziska Hecker was funded by the Bayer AG LSC initiative.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2025.02.020.

References

[1] Chandrasekaran SN, Ceulemans H, Boyd JD, Carpenter AE. Image-based profiling for drug discovery: due for a machine-learning upgrade? Nat Rev Drug Discov 2021;20:145–59. https://doi.org/10.1038/s41573-020-00117-w.

[2] Chen ZS, Pham C, Doron M, Wang S, Moshkov N, Plummer BA, et al. CHAMMI: A benchmark for channel-adaptive models in microscopy imaging. Adv Neural Inf Process Syst 2023.

[3] Xun D, Wang R, Zhang X, Wang Y. Microsnoop: a generalist tool for microscopy image representation. Innovation 2024;5. https://doi.org/10.1016/j.xinn.2023.100541.

[4] Bray M-A, Singh S, Han H, Davis CT, Borgeson B, Hartland C, et al. Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. Nat Protoc 2016;11:1757–74. https://doi.org/10.1038/nprot.2016.105.

[5] Cimini BA, Chandrasekaran SN, Kost-Alimova M, Miller L, Goodale A, Fritchman B, et al. Optimizing the cell painting assay for image-based profiling. Nat Protoc 2023;18:1981–2013. https://doi.org/10.1038/s41596-023-00840-9.

[6] Ando D.M., McLean C.Y., Berndl M. Improving Phenotypic Measurements in High-Content Imaging Screens, bioRxiv; 2017. https://doi.org/10.1101/161422.

[7] Kensert A, Harrison PJ, Spjuth O. Transfer learning with deep convolutional neural networks for classifying cellular morphological changes. SLAS Discov Adv Sci Drug Discov 2019;24:466–75. https://doi.org/10.1177/2472555218818756.

[8] Caicedo JC, McQuin C, Goodman A, Singh S, Carpenter AE. Weakly supervised learning of single-cell feature embeddings. Proc IEEE Conf Comput Vis Pattern Recognit 2018:9309–18. https://doi.org/10.1109/CVPR.2018.00970.

[9] Kraus O, Grys BT, Ba J, Chong Y, Frey BJ, Boone C, et al. Automated analysis of high-content microscopy data with deep learning. Mol Syst Biol 2017;13:924. https://doi.org/10.15252/msb.20177551.

[10] Moshkov N, Bornholdt M, Benoit S, Smith M, McQuin C, Goodman A, et al. Learning representations for image-based profiling of perturbations. Nat Commun 2024;15:1594. https://doi.org/10.1038/s41467-024-45999-1.

[11] Balestriero R., Ibrahim M., Sobal V., Morcos A., Shekhar S., Goldstein T., et al. A Cookbook of Self-Supervised Learning, arXiv [cs.CV]; 2023. https://doi.org/10.48550/arXiv.2304.12210.

[12] Caron M, Touvron H, Misra I, Jégou H, Mairal J, Bojanowski P, et al. Emerging properties in self-supervised vision transformers. Proc Int Conf Comput Vis 2021:9650–60. https://doi.org/10.1109/ICCV48922.2021.00951.

[13] Chen T, Kornblith S, Norouzi M, Hinton G. A simple framework for contrastive learning of visual representations. Int Conf Mach Learn 2020:1597–607.

[14] He K, Chen X, Xie S, Li Y, Dollár P, Girshick R. Masked autoencoders are scalable vision learners. Proc IEEE CVF Conf Comput Vis Pattern Recognit 2022:16000–9.

[15] Oquab M., Darcet T., Moutakanni T., Vo H., Szafraniec M., Khalidov V., et al. DINOv2: Learning Robust Visual Features without Supervision, arXiv; 2023. https://doi.org/10.48550/arXiv.2304.07193.

[16] Zhou J., Wei C., Wang H., Shen W., Xie C., Yuille A., et al. iBOT: Image BERT Pre-Training with Online Tokenizer, arXiv [cs.CV]; 2022.

[17] Kim V., Adaloglou N., Osterland M., Morelli F.M., Halawa M., König T., et al. Self-supervision advances morphological profiling by unlocking powerful image representations, bioRxiv; 2024. https://doi.org/10.1101/2023.04.28.538691.

[18] Doron M., Moutakanni T., Chen Z.S., Moshkov N., Caron M., Touvron H., et al. Unbiased single-cell morphology with self-supervised vision transformers, bioRxiv; 2023. https://doi.org/10.1101/2023.06.16.545359.

[19] Cross-Zamirski J.O., Williams G., Mouchet E., Schönlieb C.-B., Turkki R., Wang Y. Self-Supervised Learning of Phenotypic Representations from Cell Images with Weak Labels, arXiv [cs.CV]; 2022. https://doi.org/10.48550/arXiv.2209.07819.

[20] Haslum JF, Matsoukas C, Leuchowius K-J, Müllers E, Smith K. Metadata-guided consistency learning for high content images. Med Imaging Deep Learn 2022:918–36. https://doi.org/10.48550/arXiv.2212.11595.

[21] Yao H, Hanslovsky P, Huetter J-C, Hoeckendorf B, Richmond D. Weakly supervised set-consistency learning improves morphological profiling of single-cell images.

IEEE CVF Conf Comput Vis Pattern Recognit Workshop 2024:6978–87. https://doi.org/10.1109/CVPRW63382.2024.00691.
[22] Kobayashi H, Cheveralls KC, Leonetti MD, Royer LA. Self-supervised deep learning encodes high-resolution features of protein subcellular localization. Nat Methods 2022;19:995–1003. https://doi.org/10.1038/s41592-022-01541-z.
[23] Kraus O, Kenyon-Dean K, Saberian S, Fallah M, McLean P, Leung J, et al. Masked Autoencoders are Scalable Learners of Cellular Morphology. Proc. IEEECVF Conf. Comput. Vis. Pattern Recognit. 2024. p. 11757–68.
[24] Perakis A, Gorji A, Jain S, Chaitanya K, Rizza S, Konukoglu E. Contrastive learning of single-cell phenotypic representations for treatment classification. Mach Learn Med Imaging 2021;12966:565–75. https://doi.org/10.1007/978-3-030-87589-3_58.
[25] Dosovitskiy A., Beyer L., Kolesnikov A., Weissenborn D., Zhai X., Unterthiner T., et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, arXiv [cs.CV]; 2021. https://doi.org/10.48550/arXiv.2010.11929.
[26] Bao Y, Sivanandan S, Karaletsos T. Channel vision transformers: an image is worth C x 16 x 16 words. Int Conf Learn Represent ICLR 2024. https://doi.org/10.48550/arXiv.2309.16108.
[27] Bourriez N, Bendidi I, Cohen E, Watkinson G, Sanchez M, Bollot G, et al. ChAda-ViT: channel adaptive attention for joint representation learning of heterogeneous microscopy images. Proc IEEECVF Conf Comput Vis Pattern Recognit 2024: 11556–65.
[28] Pham C., Plummer B.A. Enhancing Feature Diversity Boosts Channel-Adaptive Vision Transformers, arXiv [cs.CV]; 2024.
[29] Hua SBZ, Lu AX, Moses AM. CytoImageNet: a large-scale pretraining dataset for bioimage transfer learning. NeurIPS Learn Mean Represent Life Workshop 2021. https://doi.org/10.48550/arXiv.2111.11646.
[30] Chandrasekaran S.N., Ackerman J., Alix E., Ando D.M., Arevalo J., Bennion M., et al. JUMP Cell Painting dataset: morphological impact of 136,000 chemical and genetic perturbations, bioRxiv; 2023. https://doi.org/10.1101/2023.03.23.534023.
[31] Rohban MH, Singh S, Wu X, Berthet JB, Bray M-A, Shrestha Y, et al. Systematic morphological profiling of human gene and allele function via Cell Painting. eLife 2017;6:e24060. https://doi.org/10.7554/eLife.24060.
[32] Caie PD, Walls RE, Ingleston-Orme A, Daya S, Houslay T, Eagle R, et al. High-content phenotypic profiling of drug response signatures across distinct cancer cells. Mol Cancer Ther 2010;9:1913–26. https://doi.org/10.1158/1535-7163.MCT-09-1148.
[33] Le T, Winsnes CF, Axelsson U, Xu H, Mohanakrishnan Kaimal J, Mahdessian D, et al. Analysis of the human protein atlas weakly supervised single-cell classification competition. Nat Methods 2022;19:1221–9. https://doi.org/10.1038/s41592-022-01606-z.
[34] Hecker FA, Leggio B, König T, Kim V, Osterland M, Gnutt D, et al. Cell Painting unravels insecticidal modes of action on *Spodoptera frugiperda* insect cells. Pest Biochem Physiol 2024;203:105983. https://doi.org/10.1016/j.pestbp.2024.105983.
[35] Way GP, Kost-Alimova M, Shibue T, Harrington WF, Gill S, Piccioni F, et al. Predicting cell health phenotypes using image-based morphology profiling. Mol Biol Cell 2021;32:995–1005. https://doi.org/10.1091/mbc.E20-12-0784.
[36] Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, Friman O, et al. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol 2006;7:R100. https://doi.org/10.1186/gb-2006-7-10-r100.
[37] Ljosa V, Caie PD, Ter Horst R, Sokolnicki KL, Jenkins EL, Daya S, et al. Comparison of methods for image-based profiling of cellular morphological responses to small-molecule treatment. SLAS Discov 2013;18:1321–9. https://doi.org/10.1177/1087057113503553.
[38] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. J Mach Learn Res 2011;12:2825–30.
[39] McInnes L., Healy J., Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction, arXiv [stat.ML]; 2020. https://doi.org/10.48550/arXiv.1802.03426.
[40] Chow YL, Singh S, Carpenter AE, Way GP. Predicting drug polypharmacology from cell morphology readouts using variational autoencoder latent space arithmetic. PLOS Comput Biol 2022;18:e1009888. https://doi.org/10.1371/journal.pcbi.1009888.
[41] Ziegler S, Sievers S, Waldmann H. Morphological profiling of small molecules. Cell Chem Biol 2021;28:300–19. https://doi.org/10.1016/j.chembiol.2021.02.012.
[42] Arevalo J, Su E, Ewald JD, van Dijk R, Carpenter AE, Singh S. Evaluating batch correction methods for image-based cell profiling. Nat Commun 2024;15:6516. https://doi.org/10.1038/s41467-024-50613-5.
[43] Lu Y, Chen J, Xiao M, Li W, Miller DD. An overview of tubulin inhibitors that interact with the colchicine binding site. Pharm Res 2012;29:2943–71. https://doi.org/10.1007/s11095-012-0828-z.
[44] Kenyon-Dean K, Wang ZJ, Urbanik J, Donhauser K, Hartford J, Saberian S, et al. ViTally consistent: scaling biological representation learning for cell microscopy. NIPS Found Models Sci Workshop 2024. https://doi.org/10.48550/arXiv.2411.02572.
[45] Quan H, Li X, Chen W, Bai Q, Zou M, Yang R, et al. Global contrast-masked autoencoders are powerful pathological representation learners. Pattern Recognit 2024;156:110745. https://doi.org/10.1016/j.patcog.2024.110745.
[46] Yang P, Yin X, Lu H, Hu Z, Zhang X, Jiang R, et al. CS-CO: a hybrid self-supervised visual representation learning method for H&E-stained histopathological images. Med Image Anal 2022;81:102539. https://doi.org/10.1016/j.media.2022.102539.