IMPERIAL

CellCLIP – Learning Perturbation Effects in Cell Painting via Text-Guided Contrastive learning

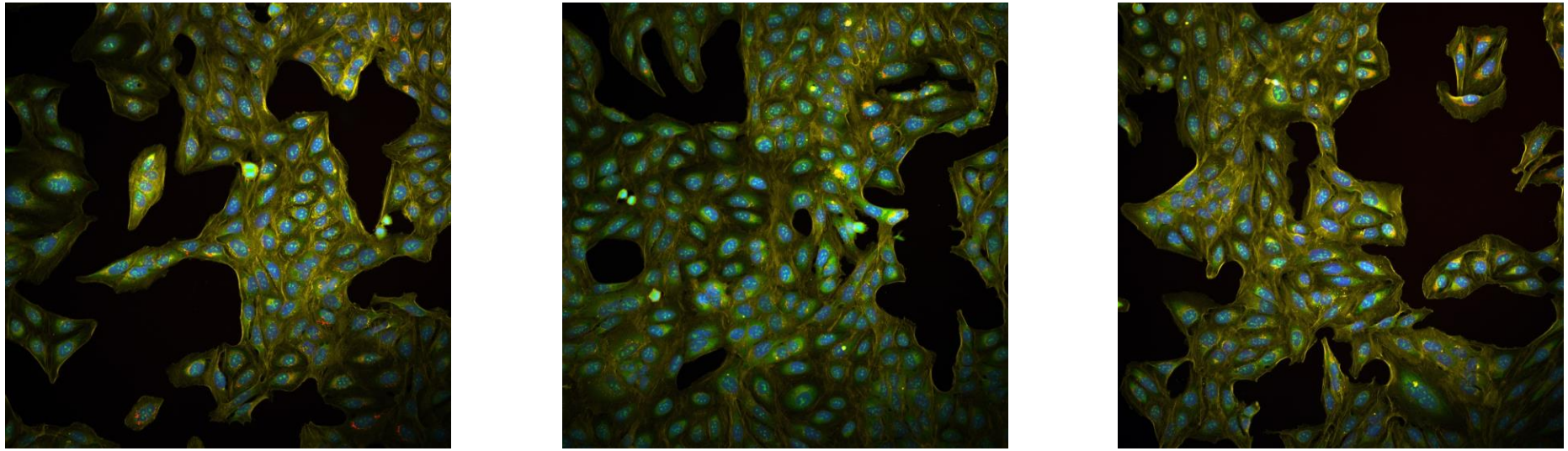
Mingyu Lu, Ethan Weinberger, Chanwoo Kim, Su-In Lee
University of Washington
NeurIPS 2025

13/11/2025 | Justin Wang

What is Cell Painting?

Cell Painting is a high-content image-based assay that uses fluorescent dyes to stain and visualize multiple cellular components, capturing **morphological information** of cells.

Using microscope, cell painting images are captured in different field of views (FOVs) of each well in a plate:



cpg0016-jump, JCP2022_085227, Site 1, 3, and 5

Why do We Need Cell Painting?

By comprehensively reflecting cell phenotype, cell painting images could **capture cellular responses to various perturbations** (small-molecule compounds, RNA interference, gene overexpression, viral infection and etc.)

Those responses could assist potential **drug discovery** by:

1. Predict the mechanism of action (MoA) of compounds.
2. Predict the assay activity of compounds.
3. Determine structure-activity relationship to guide chemistry.
4. Predict compound toxicity.
5. Profile the effect of compound mixture.
6. Advance understanding in gene and viral disease.

"Cell Painting: a decade of discovery and innovation in cellular imaging" (2024)

How do We Use Cell Painting?

Phenotypic profiling: The process of using computational method to convert a cell painting image into **representation (vectors of features)** that captures cell phenotypes.

	Classical Approach (CellProfiler)	**Deep Learning Approach (CNN, ViTs, etc.)**
Features	Predefined, hand-crafted morphological descriptors, including size, shape, texture, and etc.	Biologically relevant features recognized by deep learning models.
Performance	Captures essential morphological information and achieve moderate results.	Usually outperform classical approach in downstream tasks and mitigate batch effect, but need to be cautious about training details and generalisation.
Interpretability	Each feature is clearly interpretable, but their relationships to biological meanings might be unclear.	Features from deep learning models are challenging to interpret as they are anonymous latent variables.
Cell Segmentation	Require explicit cell segmentation in the image.	Can bypass cell segmentation entirely to get embedding
Post-processing	Many features are highly correlated and thus require feature selection.	Require much less tuning, feature are normalized and batch corrected, no need to select anything.

“Progress and new challenges in image-based profiling” (2025), “Causal integration of chemical structures improves representations of microscopy images for morphological profiling” (2025)

Multimodal Machine Learning in Cell Painting

In cell painting datasets, **images are often paired with data from other modalities:**

- Tabular Metadata (cell type, experiment info, compound name, target genes)

- Information about the perturbation:
  - Chemical structure of the applied compound (SMILES, InChI, structure files, ECFP4 fingerprint)
  - CRISPR and ORF interventions (targeting genes)

- Transcriptomic Information about the cell (Bulk RNA-Seq, LINCS signature)
- Proteomic Information about the cell (nELISA, Mass spectrometry)

Therefore, the combination of these modalities may provide a **more comprehensive representation** to reflect the cell state and reveal the **underlying relationships between cells and perturbations**.

"JUMP Cell Painting dataset: morphological impact of 136,000 chemical and genetic perturbations" (2023), "Cross modality learning of cell painting and transcriptomics data improves mechanism of action clustering and bioactivity modelling" (2025)

Multimodal Machine Learning in Cell Painting: Challenges

Substantial differences in the semantics of Cell Painting images compared to natural images

- Standard RGB channels capture **related info**, but each cell painting channel capture **different info**
- Necessary to explicitly reason between information in different channels

Multiple images are associated with the same perturbation label - "many-to-one"

- Lead to large number of **false negatives** under standard CL

Not all cell painting images faithfully represent the intended perturbation effect due to technical issues

- Lead to certain **false positives** under standard CL

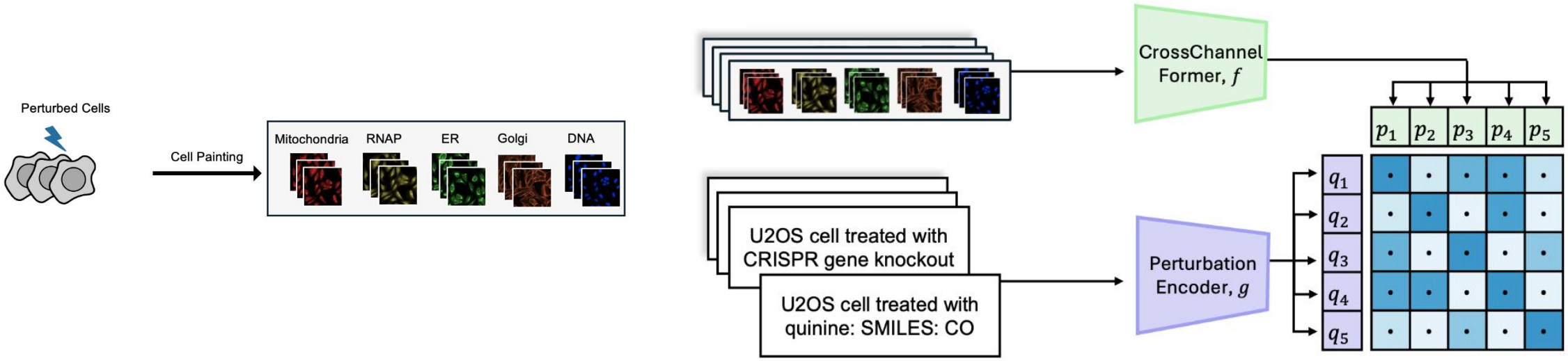
Difficulty representing different classes of perturbation in a single latent space

- Existing works mostly focus on chemical compounds (molecular fingerprint, GNN)
- Cannot be applied to CRISPR, ORF, and etc.

"CellCLIP – Learning Perturbation Effects in Cell Painting ia Text-Guided Contrastive Learning" (2025)

CellCLIP: Cross-Modal Contrastive Learning Framework for HCS Data

1. CrossChannelFormer—Image encoding scheme that learn from different channels and address false pos/neg
2. Perturbation encoding via natural language prompt that accommodates all perturbation types
3. CWCL training objectives to match similar pairs together



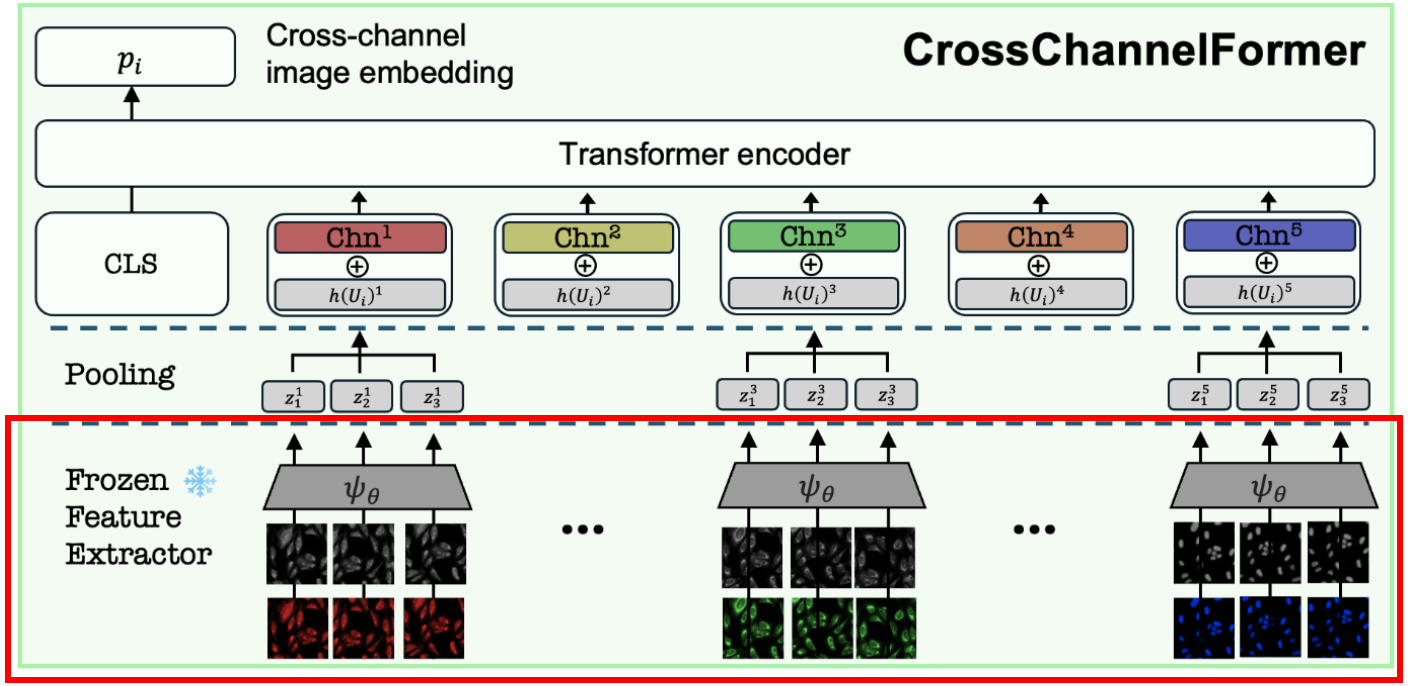
"CellCLIP – Learning Perturbation Effects in Cell Painting ia Text-Guided Contrastive Learning" (2025)

ChannelCrossFormer: The Image Encoding Scheme

Three-step process to extract image embedding that account for the characteristics of cell painting images

1. Apply pretrained DINOv2 to independent grayscale image of each cell painting channel:

$$U_i = \{u_k\}_{k=1}^{N_i}, \text{ where } u_k \in \mathbb{R}^{C\times H\times W} \qquad z_k^c = \psi_\theta(u_k^c) \in \mathbb{R}^d. \qquad z_k = [z_k^1, z_k^2, \ldots, z_k^C]^T \in \mathbb{R}^{C\times d},$$

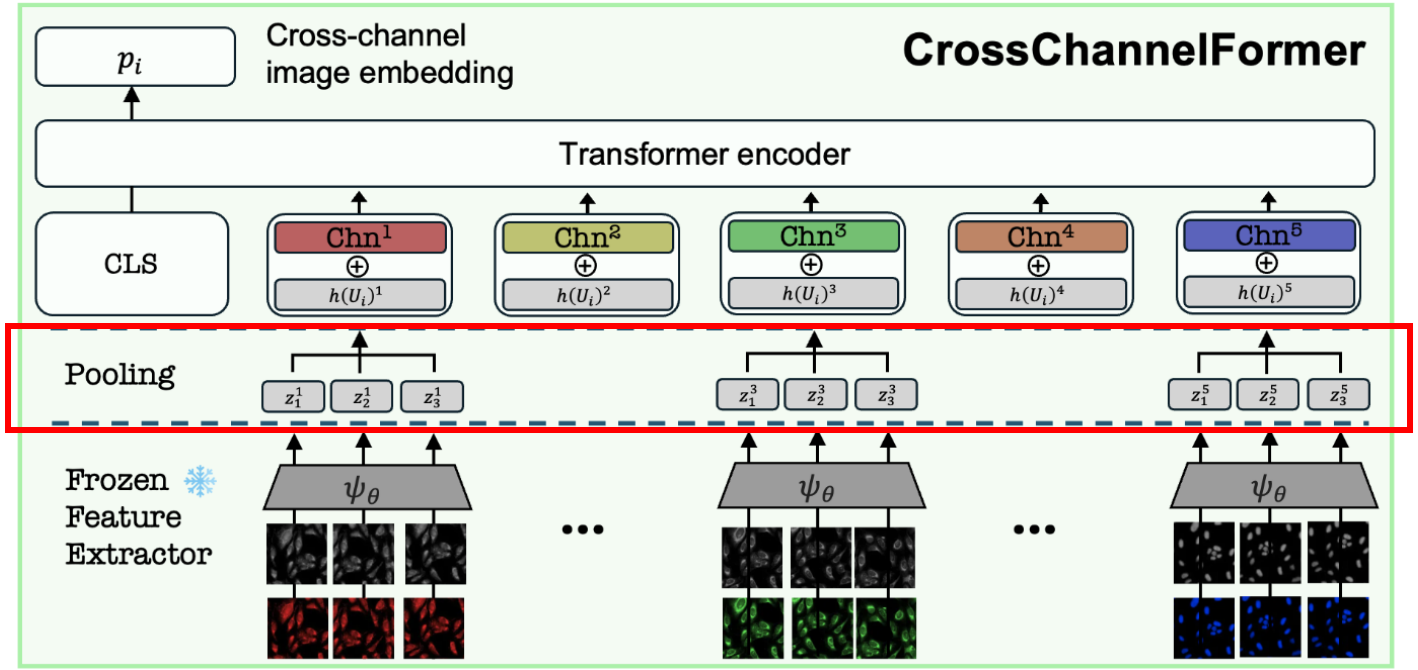


"CellCLIP – Learning Perturbation Effects in Cell Painting ia Text-Guided Contrastive Learning" (2025)

ChannelCrossFormer: The Image Encoding Scheme

Three-step process to extract image embedding that account for the characteristics of cell painting images
2. Within perturbation, pooling profiles for each channel to address false negative and false positive

$$\mu(U_i)^c = \mathcal{S}\left(\{z_k^c \mid k = 1, \ldots, N_i\}\right) \in \mathbb{R}^d,$$



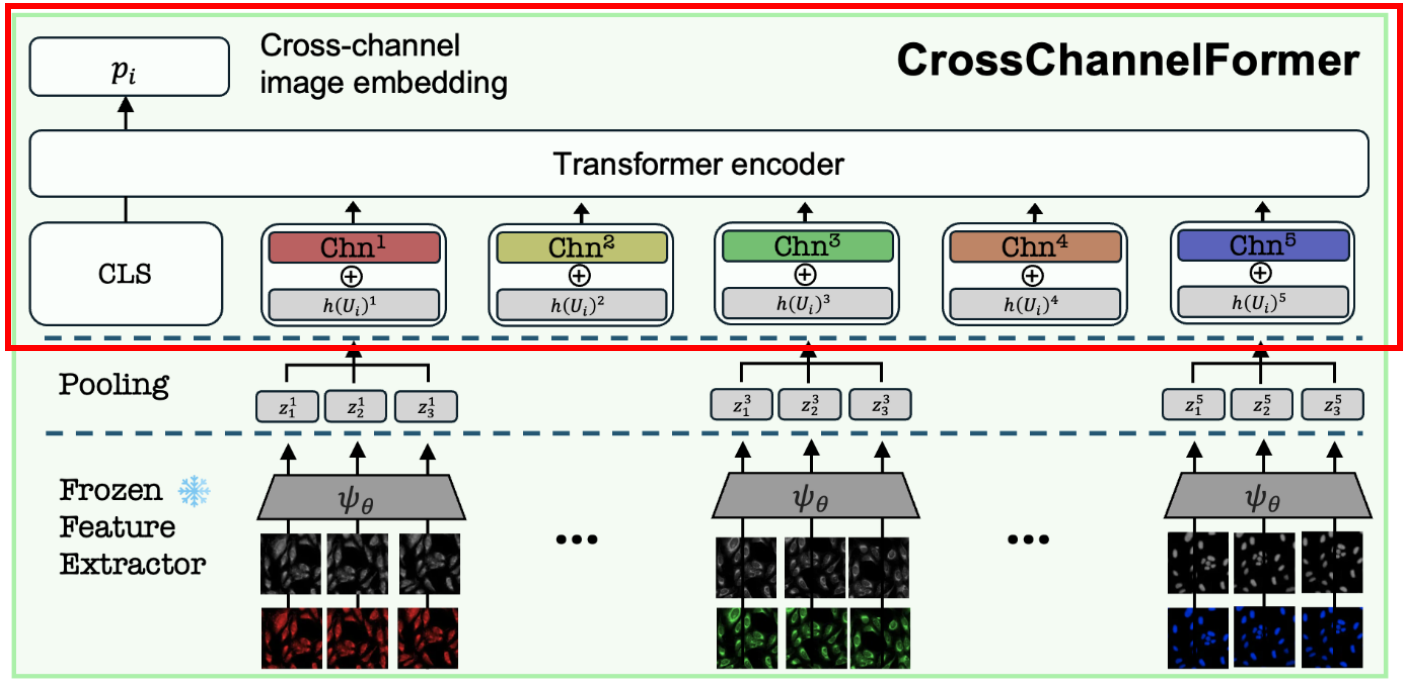
"CellCLIP – Learning Perturbation Effects in Cell Painting ia Text-Guided Contrastive Learning" (2025)

ChannelCrossFormer: The Image Encoding Scheme

Three-step process to extract image embedding that account for the characteristics of cell painting images

3. Have each input token captures the global cellular feature of specific stains and feed to Transformer

$$[\texttt{cls}, \mu(U_i)^1 + \texttt{chn}^1, \mu(U_i)^2 + \texttt{chn}^2, \ldots, \mu(U_i)^C + \texttt{chn}^C],$$



"CellCLIP – Learning Perturbation Effects in Cell Painting ia Text-Guided Contrastive Learning" (2025)

Perturbation Encoding Via Natural Language

Previous methods rely on perturbation-class-specific encoders: Morgan Fingerprint + NLP, GNN (chemical)

Motivation: representing each perturbation using descriptive text that accommodate all classes

- Description - Cell Painting Images
- Cell Types
- Perturbation Specific Details

"A cell painting image of U2OS cells treated with butyric acid, SMILES: CCCC(O)=O."

"A cell painting image of U2OS cells treated with CRISPR, targeting genes: AP2S1."

Encoder Architecture: pretrained BERT model

"CellCLIP – Learning Perturbation Effects in Cell Painting ia Text-Guided Contrastive Learning" (2025)

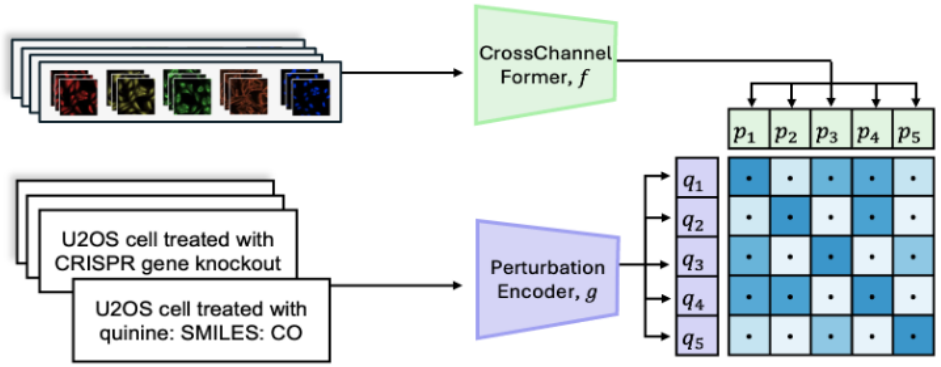
CellCLIP Training Objective

In Cell Painting, matching related perturbation are also important.

- Treating all non-matching pairs as hard negatives can degrade retrieval performance

Continuously Weighted Contrastive Loss (CWCL)

- Replace binary labels with continuous similarity-based weights
- Compute target similarities between per-perturbation pooled profiles (avg channel-wise cosine similarity)
- One Direction: profile->perturbation use CWCL, perturbation->profile use CLIP
  - Similarity in profile is meaningful, Similarity in perturbation label is not



$$\mathcal{L}_{\text{CWCL},\mathcal{U}\to\mathcal{V}} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sum_{j\in[N]} w_{ij}^{\mathcal{U}}}\left[\sum_{j=1}^{N} w_{ij}^{\mathcal{U}}\cdot\log\frac{\exp(\langle p_i\cdot q_j\rangle/\tau)}{\sum_{k=1}^{N}\exp(\langle p_i\cdot q_k\rangle/\tau)}\right]$$

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{CWCL},\mathcal{U}\to\mathcal{V}} + \mathcal{L}_{\text{CLIP},\mathcal{V}\to\mathcal{U}}.$$

"CellCLIP – Learning Perturbation Effects in Cell Painting ia Text-Guided Contrastive Learning" (2025), "CWCL: Cross-Modal Transfer with Continuously Weighted Contrastive Loss" (2023)

CellCLIP: The Experiment

1. **Cross-Modality Retrieval Between Profiles and Perturbations**
   - Retrieve test set perturbation given corresponding Cell Painting image profiles treated with the perturbation
   - And vice versa...
   - Common cell painting tasks in many literatures
   - Evaluation Metric: Recall@k, measure whether the correct perturbation exist in top-k retrieved results

2. **Evaluate Profile Embedding on Intra-modal Biologically Meaningful Tasks**
   - Replicate Detection: Distinguish replicates across batches from negative controls
   - Sister Perturbation Matching: Match cells treated with "sister" perturbations targeting the same gene
   - Zero-Shot Gene-Gene Relationship Recovery: Check profile embeddings on unseen data points preserve real biological relationships between genes

Datasets:

- Training: Bray et al.
- Evaluation: CPJUMP1, RxRx3, Bray et al.

"CellCLIP – Learning Perturbation Effects in Cell Painting ia Text-Guided Contrastive Learning" (2025)

Cross Modality Retrieval: Results

Model	Perturb-to-profile (%) ↑			Profile-to-perturb (%) ↑		
	R@1	R@5	R@10	R@1	R@5	R@10
CLOOME	0.27 ± 0.20	1.25 ± 0.42	2.46 ± 0.56	0.07 ± 0.09	1.44 ± 0.48	2.56 ± 0.89
CLOOME‡	0.45 ± 0.13	1.60 ± 0.12	3.04 ± 0.12	0.48 ± 0.16	1.79 ± 0.13	3.12 ± 0.25
CLOOME‡ (CLOOB)	0.37 ± 0.10	1.56 ± 0.10	2.75 ± 0.02	0.20 ± 0.03	1.76 ± 0.10	3.13 ± 0.17
MolPhenix* (S2L)	0.28 ± 0.11	1.56 ± 0.19	3.00 ± 0.26	0.42 ± 0.08	1.54 ± 0.03	3.01 ± 0.20
MolPhenix* (SigCLIP)	0.56 ± 0.12	2.78 ± 0.23	4.30 ± 0.16	0.66 ± 0.22	2.55 ± 0.05	4.34 ± 0.24
CellCLIP (Ours)	**1.18 ± 0.20**	**4.49 ± 0.06**	**7.37 ± 0.20**	**1.25 ± 0.10**	**4.82 ± 0.10**	**7.39 ± 0.23**

‡ Indicates that CLOOME uses the same image profile encoding procedure and architecture as MolPhenix*

Vision Encoder	Perturb. Encoder	Train Time (hr)	Perturb-to-profile (%) ↑			Profile-to-perturb (%) ↑		
			R@1	R@5	R@10	R@1	R@5	R@10
ResNet-101	△ + MLP	49.9	0.27 ± 0.20	1.25 ± 0.42	2.46 ± 0.56	0.07 ± 0.09	1.44 ± 0.48	2.56 ± 0.89
ResNet-101	△ + MPNN++	45.2	0.20 ± 0.06	1.30 ± 0.31	3.03 ± 0.21	0.15 ± 0.12	1.50 ± 0.41	2.78 ± 0.51
ResNet-101	□ + BERT	40.5	0.74 ± 0.07	2.67 ± 0.09	4.50 ± 0.29	**1.10 ± 0.46**	3.95 ± 0.49	5.97 ± 0.57
CA-MAE	□ + BERT	25.6	0.22 ± 0.07	1.57 ± 0.33	2.80 ± 0.05	0.19 ± 0.05	1.29 ± 0.10	2.52 ± 0.14
ChannelViT	□ + BERT	29.5	0.78 ± 0.15	3.23 ± 0.32	5.15 ± 0.52	0.69 ± 0.18	2.92 ± 0.39	5.01 ± 0.38
CrossChannelFormer†	□ + BERT	12.2	**1.16 ± 0.12**	3.98 ± 0.33	5.74 ± 0.36	1.05 ± 0.08	3.84 ± 0.43	5.90 ± 0.28
CrossChannelFormer	□ + BERT	**1.81**	**1.18 ± 0.20**	**4.49 ± 0.06**	**7.37 ± 0.20**	**1.25 ± 0.10**	**4.82 ± 0.10**	**7.39 ± 0.23**

†: Trained without profile pooling.

"CellCLIP – Learning Perturbation Effects in Cell Painting ia Text-Guided Contrastive Learning" (2025)

Intro-modal Biologically Meaningful Tasks: Results

Method	Replicate Detection (mAP) ↑	Sister Perturb. Matching (mAP) ↑		Gene-Gene Relationship Recovery (Recall) ↑				
		Within Class	Across Class	CORUM	HuMAP	Reactome	SIGNOR	StringDB
Cross-modal								
CLOOME	.575	.245	.026	.597	.679	.327	.309	.510
MolPhenix*	.531	.222	.011	.539	.599	.330	.297	.476
CellCLIP	**.663**	**.413**	**.043**	**.714**	**.778**	**.427**	**.388**	**.618**
Unimodal (self-supervised)								
OpenPhenom-S/16	.357	.219	.031	.649	.723	.418	.386	.579
Unimodal (weakly supervised)								
ViT-L/16	.513	.283	.032	.681	.758	.388	.380	.587

Findings

1. Cross-modal CL outperforms unimodal baselines (perturbation info help learn the profile embedding better)
2. CellCLIP performs the best among cross-modal approaches
3. Replace CWCL loss with standard CLIP loss leads to worse performance in Gene-Gene Relationship task

"CellCLIP – Learning Perturbation Effects in Cell Painting ia Text-Guided Contrastive Learning" (2025)

IMPERIAL

Thanks!

13/11/2025 | Justin Wang