

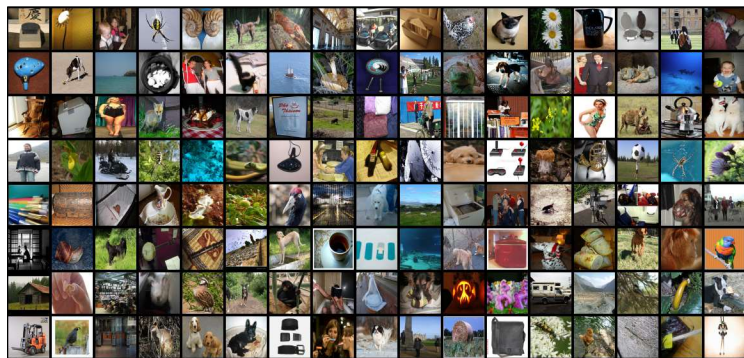
# CytoImageNet: A large-scale pretraining dataset for bioimage transfer learning

Stanley Z. Hua<sup>1</sup>, Alex X. Lu<sup>2</sup>, Alan M. Moses<sup>1</sup>

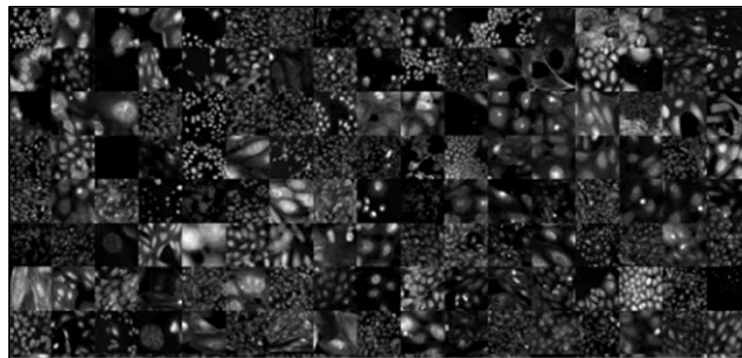
University of Toronto<sup>1</sup>, University of Toronto (Current Affiliation: Microsoft Research)<sup>2</sup>

**CytoImageNet contains 890K microscopy images and 894 classes.**

## IMAGENET-1K



## CYTOIMAGENET

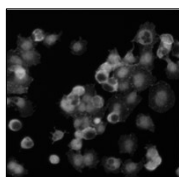


# of Images	1.3 million images	890K images
# of Classes	1000 classes	894 classes
Source of Label	Human annotated labels	Metadata labels
# of Channels	3 channel images (RGB)	1 channel images (grayscale)
Type of Images	Natural scene images	Microscopy images
Source of Images	Open-source images	Images from 40 open-source datasets
Properties	Images of varying scale	Images of varying scale (and rotation)

**Concatenating CytoImageNet-trained and ImageNet-trained features is the new state-of-the-art in bioimage transfer learning.**

### Downstream Microscopy Classification Tasks

#### BBBC021



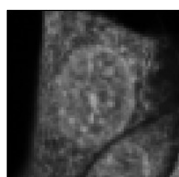
Human Cells  
Mechanism-Of-Action

#### CYCLoPs



Yeast Cells  
Subcellular Localization

#### COOS7



Mouse Cells  
Subcellular Localization

**Table 1.** kNN accuracy using EfficientNetB0 penultimate layer features.

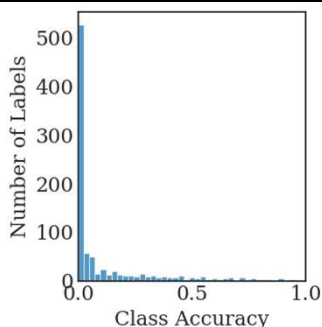
Downstream Task	Random	ImageNet	CytoImageNet	<b>Fusion</b>
BBBC021	27.18%	<b>83.50%</b>	<b>83.50%</b>	<b>86.41%</b>
CYCLoPs	53.06%	<b>68.47%</b>	65.19%	<b>77.97%</b>
COOS7 Test Set 1	65.68%	<b>88.87%</b>	88.58%	<b>94.80%</b>
COOS7 Test Set 2	67.81%	88.93%	<b>89.37%</b>	<b>95.07%</b>
COOS7 Test Set 3	48.77%	<b>75.91%</b>	65.97%	<b>78.98%</b>
COOS7 Test Set 4	51.88%	<b>82.19%</b>	78.52%	<b>87.47%</b>

## Data Availability



<https://www.kaggle.com/stanleyhua/cytoimagenet>

## CytoImageNet Pretraining yields useful features despite Poor Accuracy on the Validation Set.



**Table 2.** EfficientNetB0 Validation Accuracy

Dataset	Top-1 Validation Accuracy
ImageNet-1K [5]	77.1%
CytoImageNet	11.32%

**402** of the 894 labels had **no** correctly predicted images in the CytoImageNet validation set.

## REFERENCES

- [1] Caicedo JC, McQuin C, Goodman A, Singh S, Carpenter AE. Weakly supervised learning of single-cell feature embeddings. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018. doi:[10.1109/cvpr.2018.00970](https://doi.org/10.1109/cvpr.2018.00970)
- [2] Kraus OZ, Gryns BT, Ba J, Chong Y, Frey BJ, Boone C, Andrews BJ. Automated Analysis of high-content microscopy data with Deep Learning. Molecular Systems Biology. 2017;13(4):924.doi:[10.15252/msb.20177551](https://doi.org/10.15252/msb.20177551)
- [3] Pawlowski N, Caicedo JC, Singh S, Carpenter AE, Storkey A. Automating morphological profiling with generic deep convolutional networks. bioRxiv. 2016. doi:[10.1101/085118](https://doi.org/10.1101/085118)
- [4] Huh M, Agrawal P, Efros AA. What makes ImageNet good for transfer learning? arXiv. 2016;[1608.08614](https://arxiv.org/abs/1608.08614).
- [5] Tan M, Le Q. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In: 36th International Conference on Machine Learning (ICML 2019) Long Beach, California, USA, 9-15 June 2019. Red Hook, NY: Curran Associates, Inc; 2019.