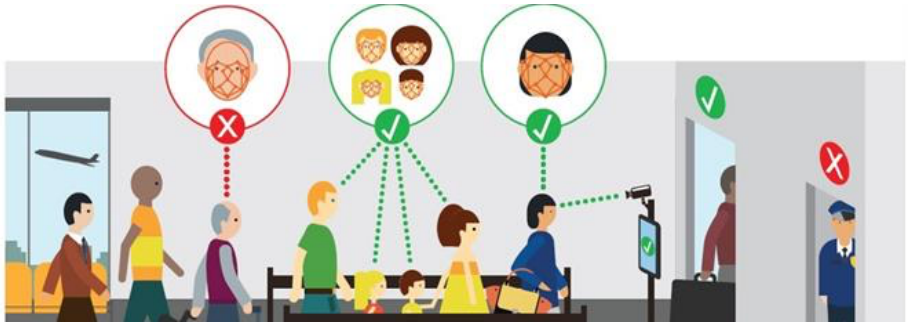
The Will of Computer Vision

Zongwei Zhou

Dept. Biomedical Informatics
Arizona State University

@MrGiovanni

Facial recognition:



Self-driving cars:

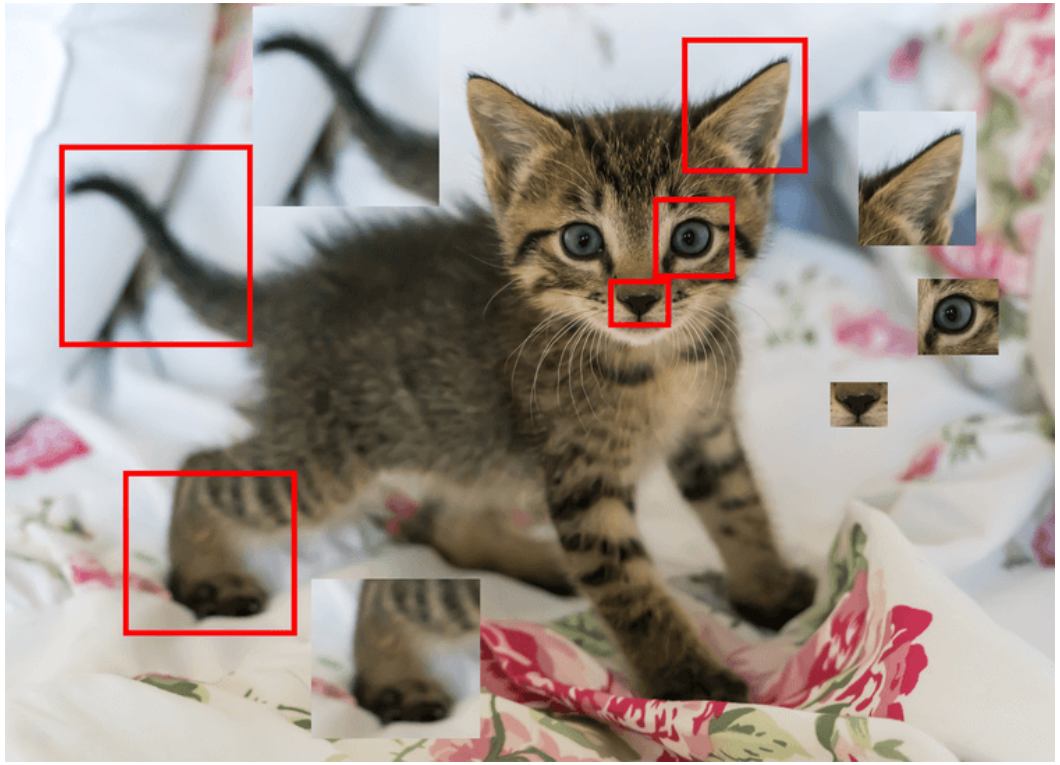
Medical imaging:

Robotic surgery:

https://www.thalesgroup.com/en/markets/digital-identity-and-security/government/biometrics/facial-recognition
https://www.fastcompany.com/90145568/19-artists-draw-their-perfect-self-driving-car
https://blogs.nvidia.com/blog/2020/08/12/lunit-insight-cxr/
https://www.cambridgeindependent.co.uk/business/milestone-as-cmr-surgical-s-versius-robot-is-used-by-nhs-hospitals-for-first-time-9100442/

The Will of Computer Vision was

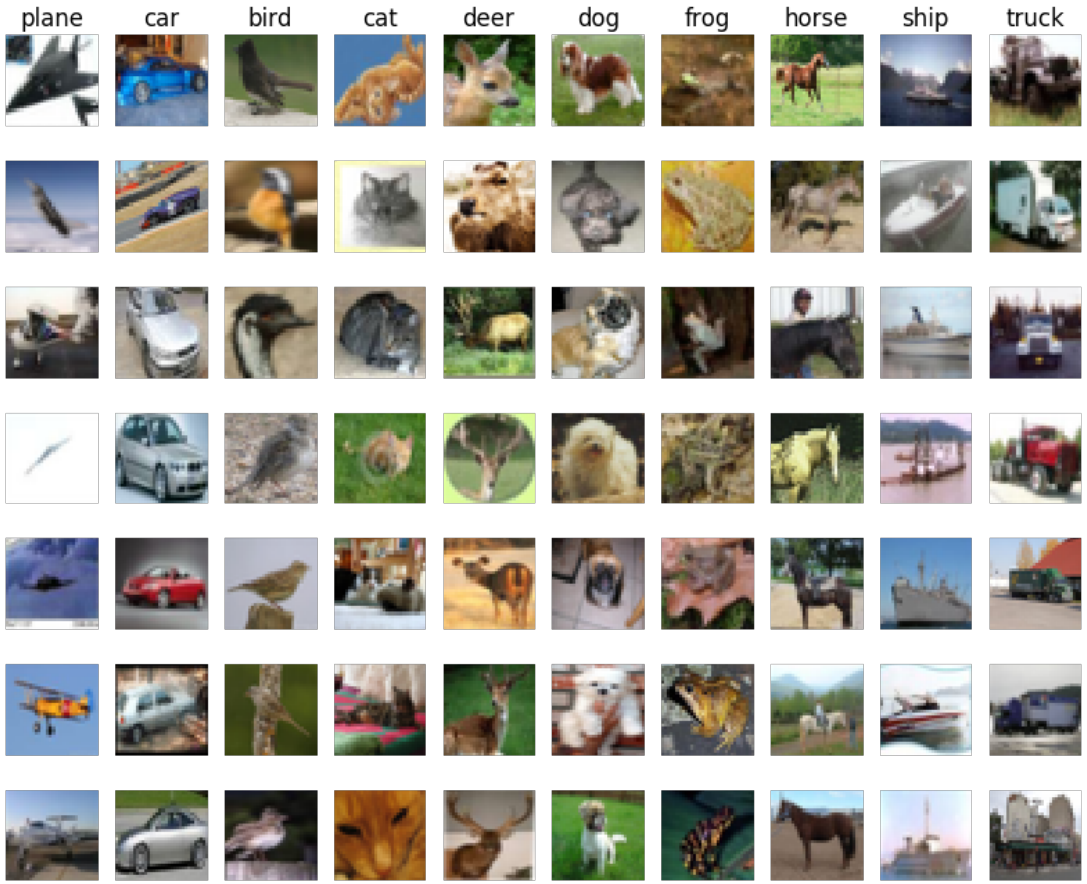
Feature Engineering and Matching



https://techxplore.com/news/2017-03-survival-ai-revolution.html

The Will of Computer Vision was

Large-scale Image Categorization



Categorization vs. Contrast

Categories in the real world are

Non-orthogonal

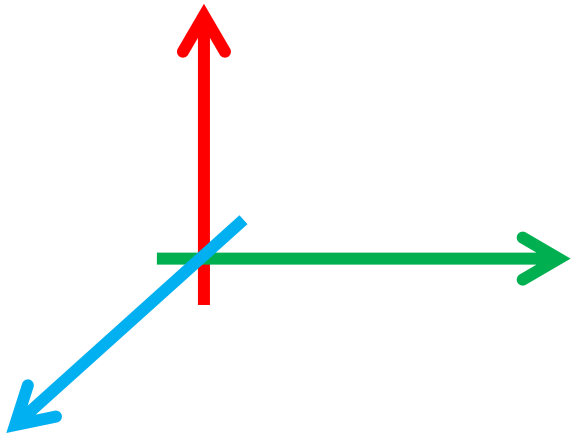
Unbalanced

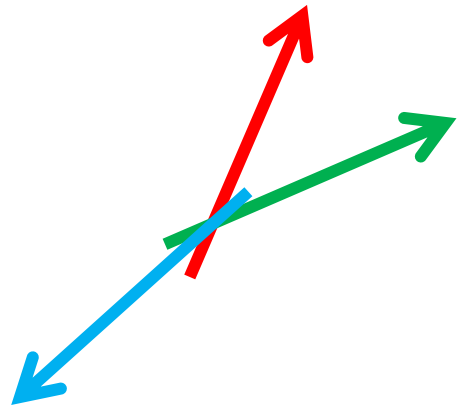
Exponential

(1,0,0)

(0,1,0)

(0,0,1)





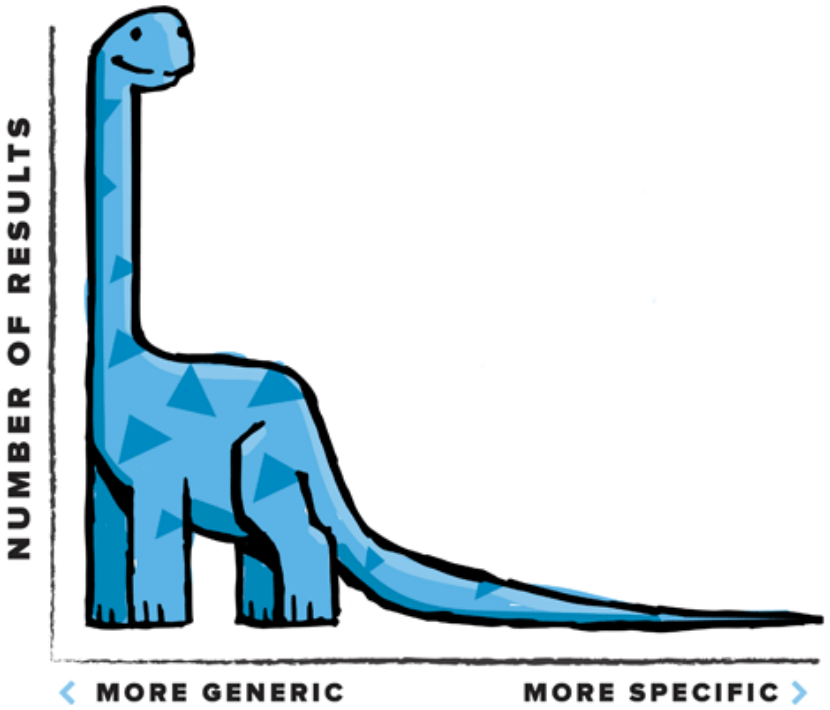
Categorization vs. Contrast

Categories in the real world are

Non-orthogonal

Unbalanced

Exponential



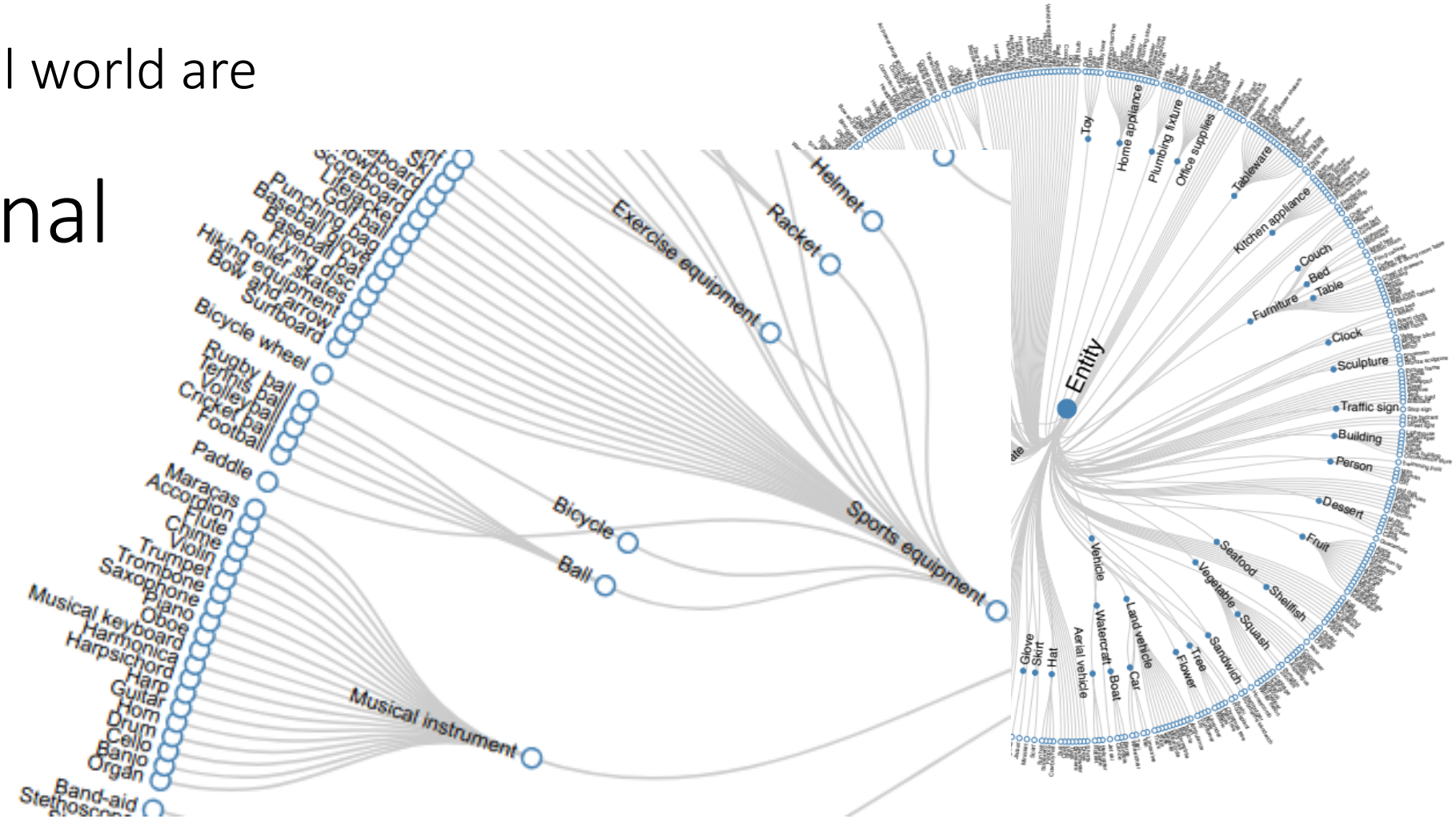
Categorization vs. Contrast

Categories in the real world are

Non-orthogonal

Unbalanced

Exponential



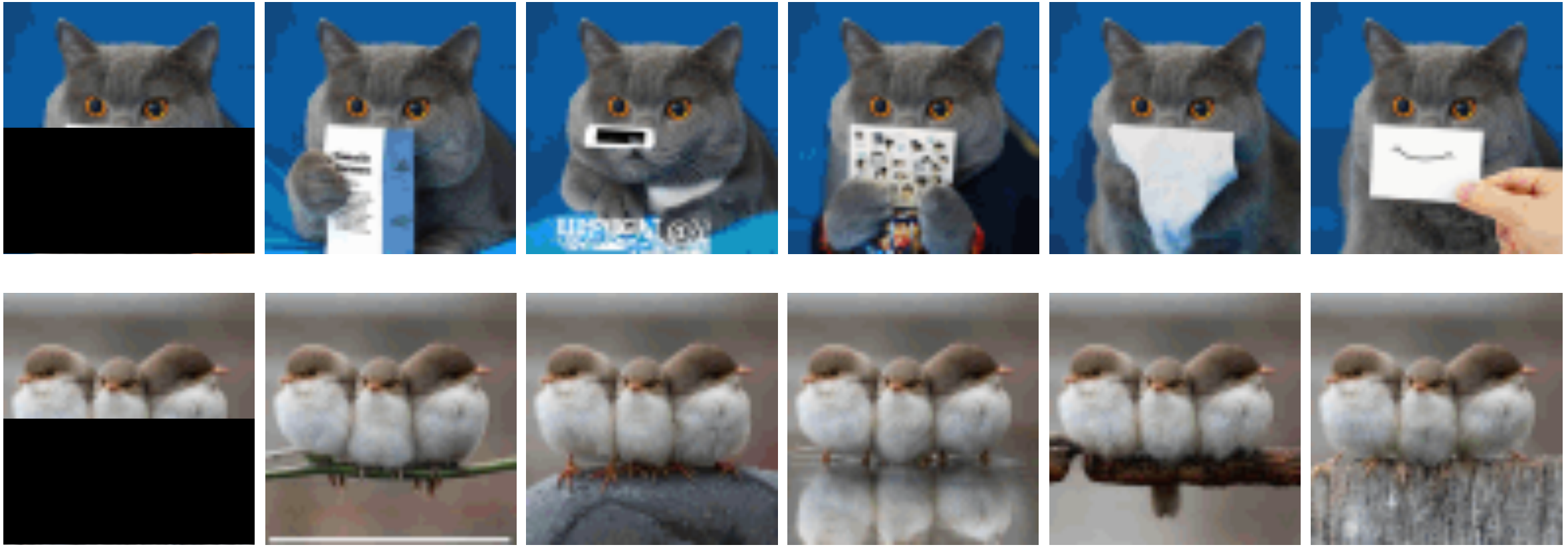
The Will of Computer Vision was

Large-scale Hidden Context Prediction

THE WILL [redacted] MPUTER VISION

The Will of Computer Vision was

Large-scale Hidden Context Prediction



https://openai.com/blog/image-gpt/

The Will of Computer Vision was

Large-scale Multiple View Discrimination

The Will of Computer Vision was

Large-scale Multiple View Discrimination

Feng, R., Zhou, Z., Gotway, M.B. and Liang, J., 2020. Parts2Whole: Self-supervised Contrastive Learning via Reconstruction. In Domain Adaptation and Representation Transfer, and Distributed and Collaborative Learning (pp. 85-95). Springer, Cham.

WILL: A Novel Pre-training

So Many Pre-trainings: A Survey

	ISIC			ChestX		
	5-shot	20-shot	50-shot	5-shot	20-shot	50-shot
InsDis	43.90 ± 0.55	52.19 ± 0.53	55.76 ± 0.50	25.67 ± 0.42	29.13 ± 0.44	31.77 ± 0.44
MoCo-v1	**44.42 ± 0.55**	**53.79 ± 0.54**	56.81 ± 0.52	25.92 ± 0.45	30.00 ± 0.43	32.74 ± 0.43
PCL-v1	33.21 ± 0.48	38.01 ± 0.44	39.77 ± 0.45	23.33 ± 0.40	25.54 ± 0.43	27.40 ± 0.42
PIRL	43.89 ± 0.54	53.24 ± 0.56	56.89 ± 0.52	25.60 ± 0.41	29.48 ± 0.45	31.44 ± 0.47
PCL-v2	37.47 ± 0.52	44.40 ± 0.52	46.82 ± 0.46	24.87 ± 0.42	28.28 ± 0.42	30.56 ± 0.43
SimCLR-v1	43.99 ± 0.55	53.00 ± 0.54	56.16 ± 0.53	26.36 ± 0.44	30.82 ± 0.43	33.16 ± 0.47
MoCo-v2	42.60 ± 0.55	52.39 ± 0.49	55.68 ± 0.53	25.26 ± 0.44	29.43 ± 0.45	32.20 ± 0.43
SimCLR-v2	43.66 ± 0.58	53.15 ± 0.53	56.83 ± 0.54	26.34 ± 0.44	30.90 ± 0.44	33.23 ± 0.47
SeLa-v2	39.97 ± 0.55	48.43 ± 0.54	51.31 ± 0.52	25.60 ± 0.44	30.43 ± 0.46	32.81 ± 0.44
InfoMin	39.03 ± 0.55	48.21 ± 0.54	51.58 ± 0.51	25.78 ± 0.44	29.48 ± 0.44	31.58 ± 0.44
BYOL	43.09 ± 0.56	53.76 ± 0.55	**58.03 ± 0.52**	26.39 ± 0.43	30.71 ± 0.47	**34.17 ± 0.45**
DeepCluster-v2	40.73 ± 0.59	49.91 ± 0.53	53.65 ± 0.54	26.51 ± 0.45	**31.51 ± 0.45**	**34.17 ± 0.48**
SwAV	39.66 ± 0.54	47.08 ± 0.50	51.10 ± 0.50	**26.54 ± 0.48**	30.91 ± 0.45	33.86 ± 0.46
Supervised	39.38 ± 0.58	48.79 ± 0.53	52.54 ± 0.56	25.22 ± 0.41	29.26 ± 0.44	32.34 ± 0.45

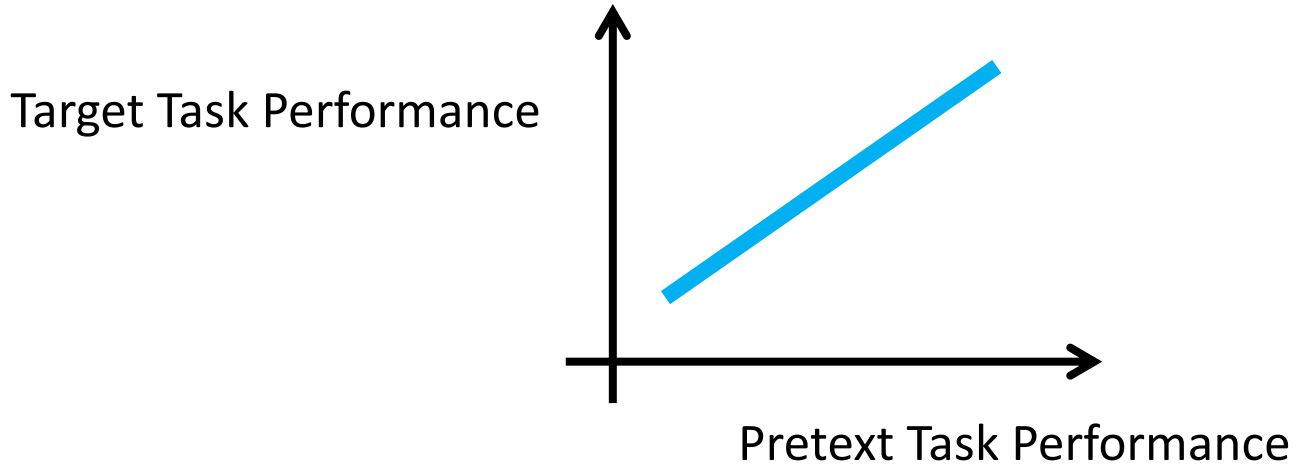
Ericsson, L., Gouk, H. and Hospedales, T.M., 2020. How Well Do Self-Supervised Models Transfer?. arXiv preprint arXiv:2011.13377./

The Will of Computer Vision—Literature

WILL: A Novel Pre-training

So Many Pre-trainings: A Survey

Do Better WILL Transfer?



Kornblith, S., Shlens, J. and Le, Q.V., 2019. Do better imagenet models transfer better?. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 2661-2671).

“Do not Define Anything”

The choice of augmented views

The choice of model architectures

The choice of pretext tasks

The Will of Computer Vision

The Will of Computer Vision

of Computer The Will Vision

Vision of The Computer Will

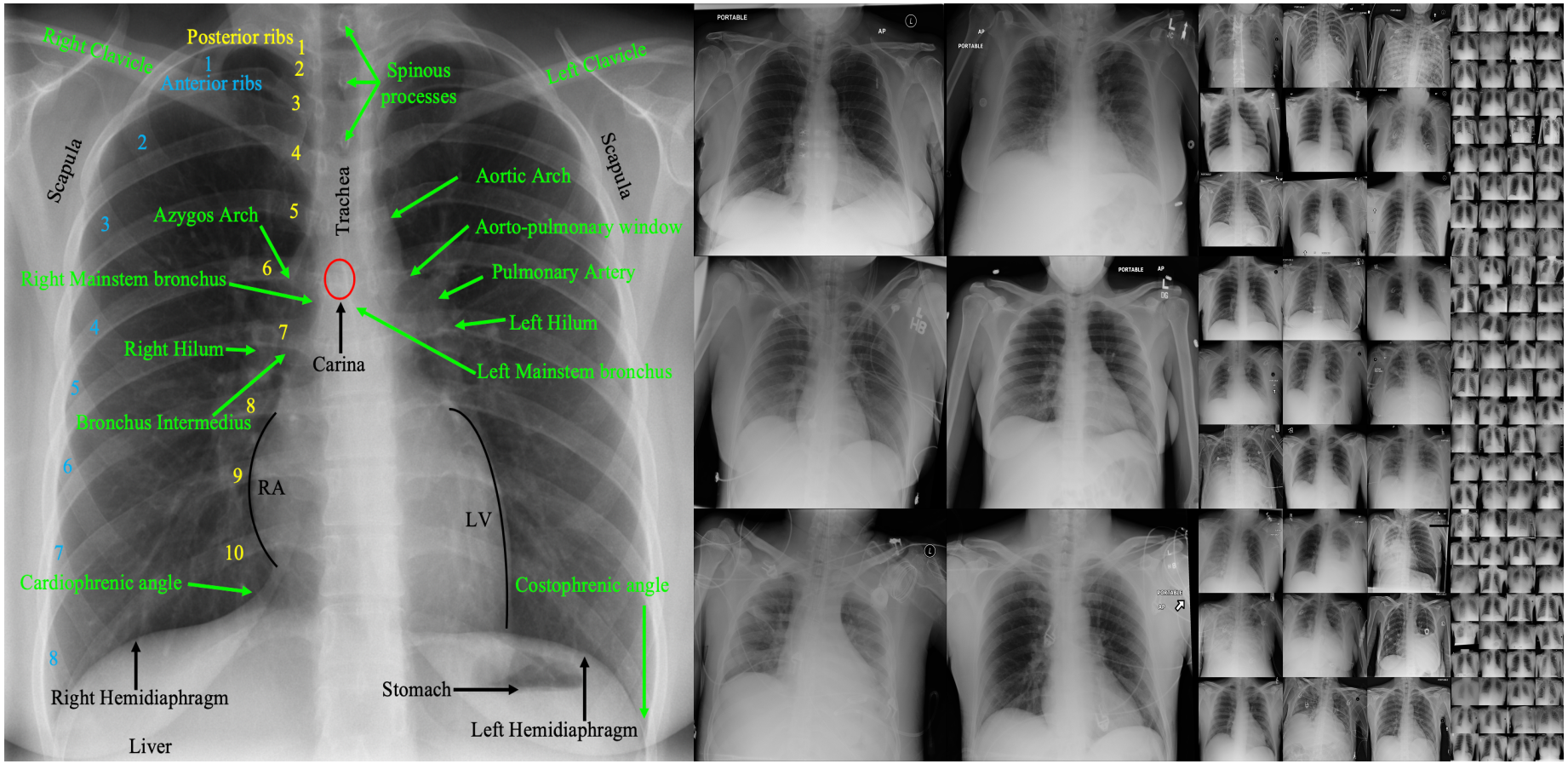
Will of Computer Vision The

Vision The Computer Will of

The Vision of Computer Will

.....

Medical Images Contain Consistent Anatomical Structures



Haghighi, F., Taher, M.R.H., Zhou, Z., Gotway, M.B. and Liang, J., 2020, October. Learning Semantics-enriched Representation via Self-discovery, Self-classification, and Self-restoration. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 137-147). Springer, Cham.

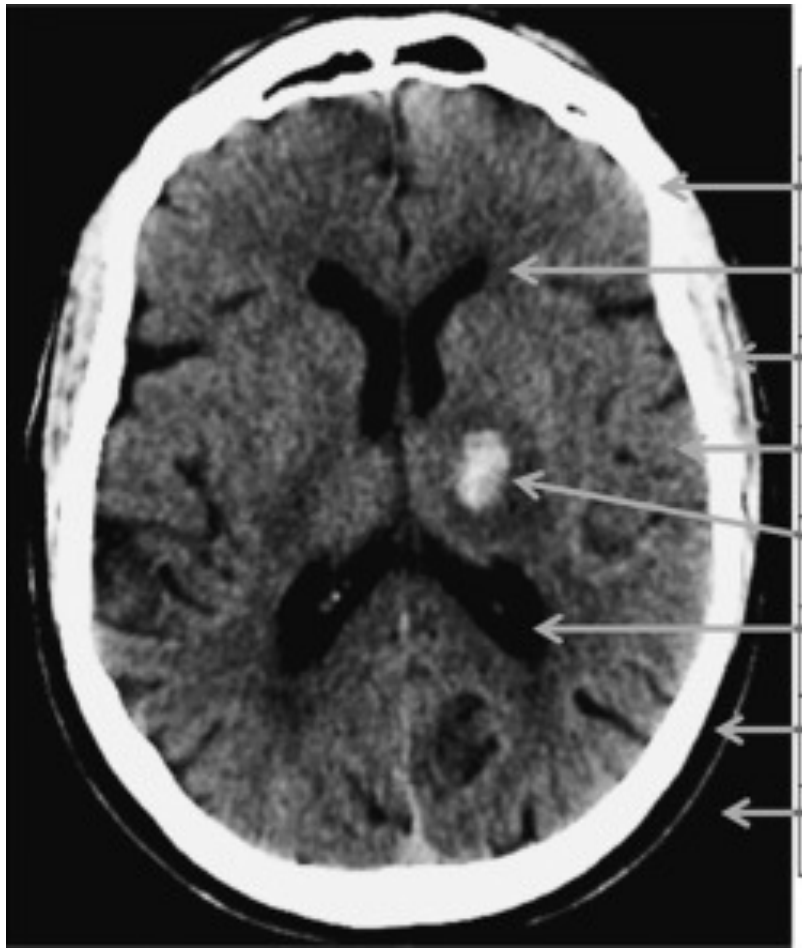
Haghighi, F., Taher, M.R.H., Zhou, Z., Gotway, M.B. and Liang, J., 2021, Transferable Visual Word. IEEE Transactions on Medical Imaging. (coming soon)

Natural Images *Sometimes* Contain Consistent Structures

https://www.biometricupdate.com/202001/facial-recognition-datasets-and-controversies-drive-biometrics-and-digital-id-news-of-the-week

Images are the Language of the Creator

Medical Images Convey Physical Meaning



CT Number Ranges in Hounsfield Units (HU)	
Bone	+1000
White matter	+20 to 30
Muscle	+20 to 40
Gray matter	+30 to 40
Hemorrhage	+65 to +95
CSF (water)	0
Fat	-30 to -70
Air	-1000

https://www.sciencedirect.com/topics/medicine-and-dentistry/hounsfield-scale
Zhou, Z., Sodha, V., Pang, J., Gotway, M.B. and Liang, J., 2020. Models genesis. Medical image analysis, 67, p.101840.

Medical Images are High Dimensional

http://henrybetts.co.uk/an-attempt-at-bullet-time/

The Will of Computer Vision

Pretext tasks and objective functions—

- Identifying the descriptive attributes from the image
- Minimizing the error between computer predictions and human labels
- Predicting some hidden portion of the image
- Distinguishing different views of context and instance
- **Learning characteristics of special image modality**