Graph Your Own Prompt

Xi Ding, Lei Wang, Piotr Koniusz, Yongsheng Gao

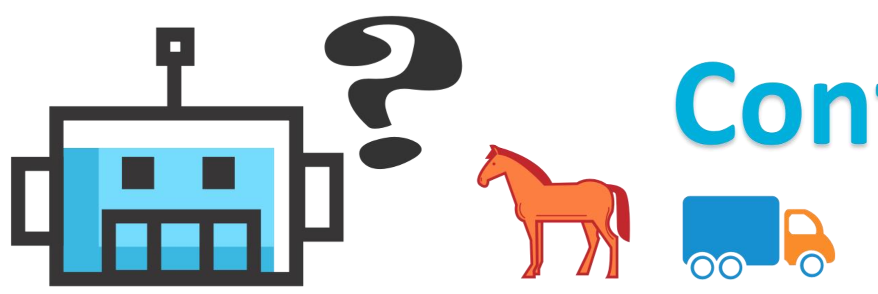
Griffith University, Queensland, Australia
CSIRO
Data61
Australian National University
UNSW Sydney

Code

Paper

Motivation

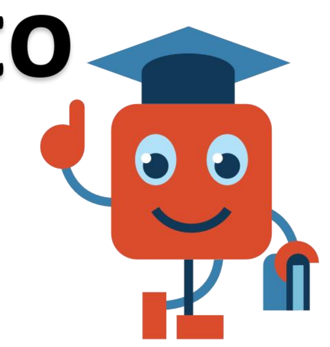
- ✓ Deep networks learn rich features, but these features often do not match semantic class structure.
- ✓ Samples predicted as the same class may still appear far apart in feature space, hurting generalization.



Confused in abstract space

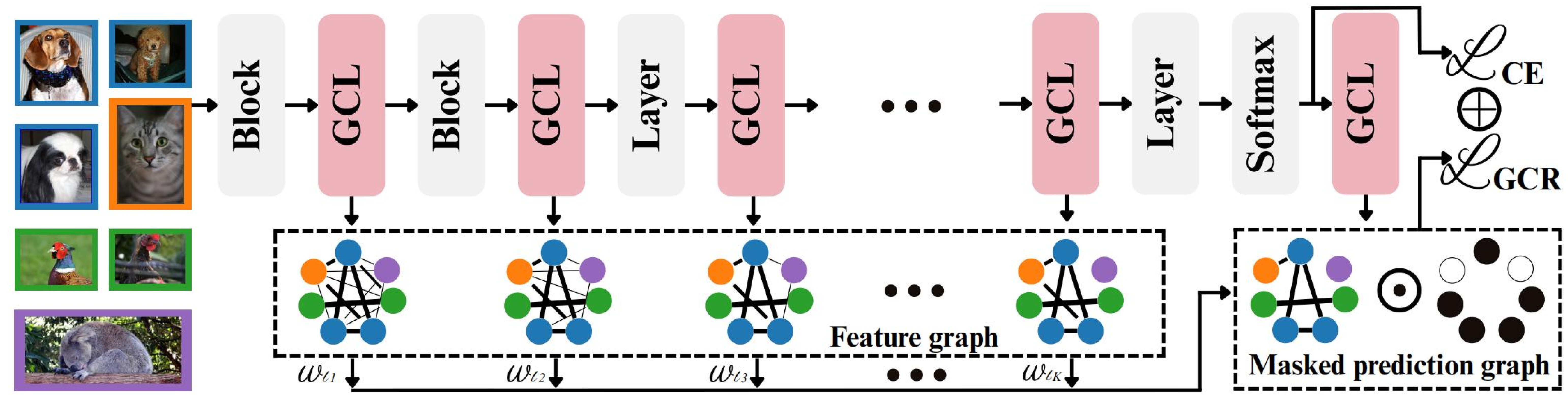
Four legs? Hmm... A car? Or a horse?

Why not use your own predictions to refine and clean feature structure?



Strength

Lightweight **Model-agnostic** **Parameter-free** **Portable**



Self-prompting: The model learns from its own outputs, reinforcing semantic structure

Method

We use cosine similarity with non-negative values:

$$F_{ij}^{(l)} = \mathrm{ReLU}(\cos(x_i^{(l)}, x_j^{(l)})), \quad i, j = 1, \ldots, n. \tag{1}$$

From the prediction logits $Z = [z_1^\top, \ldots, z_n^\top]^\top$ of the same batch:

- apply softmax to obtain class probability vectors $p_i = \mathrm{softmax}(z_i)$,
- compute pairwise cosine similarity between prediction vectors:

$$S_{ij} = \mathrm{ReLU}(\cos(p_i, p_j)). \tag{2}$$

To focus on reliable semantic relations, we build a binary mask $M \in \{0,1\}^{n \times n}$:

$$M_{ij} = \begin{cases} 1, & \text{if } y_i = y_j, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

The masked prediction graph $P \in \mathbb{R}^{n \times n}$ is then

$$P_{ij} = M_{ij} \odot S_{ij}, \tag{4}$$

where $\odot$ denotes elementwise multiplication.

The layer-wise **graph consistency loss** is

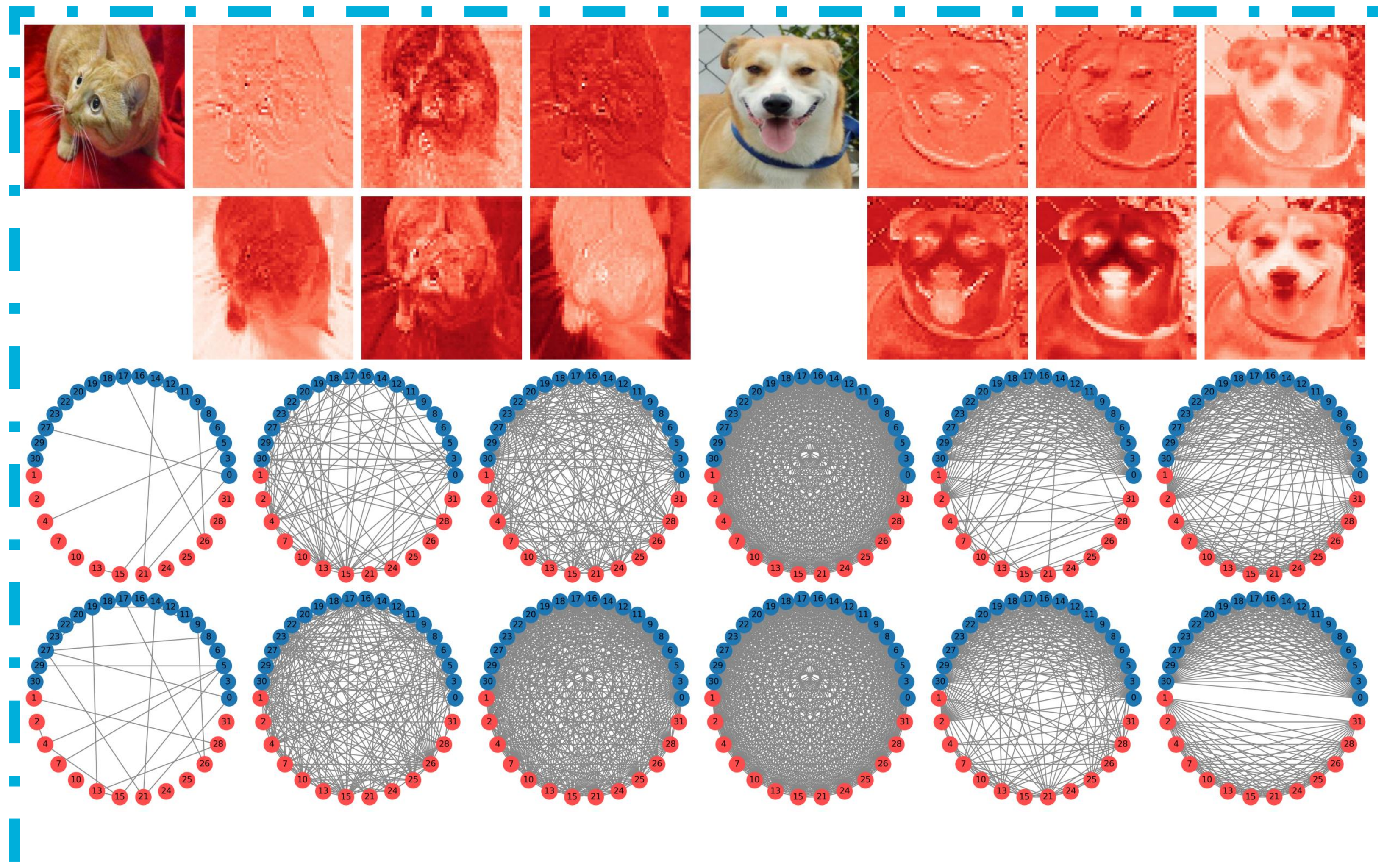
$$\mathcal{L}_{\mathrm{GCR}}^{(l)} = \|\mathrm{triu}(F^{(l)}) - \mathrm{triu}(P)\|_F^2. \tag{5}$$

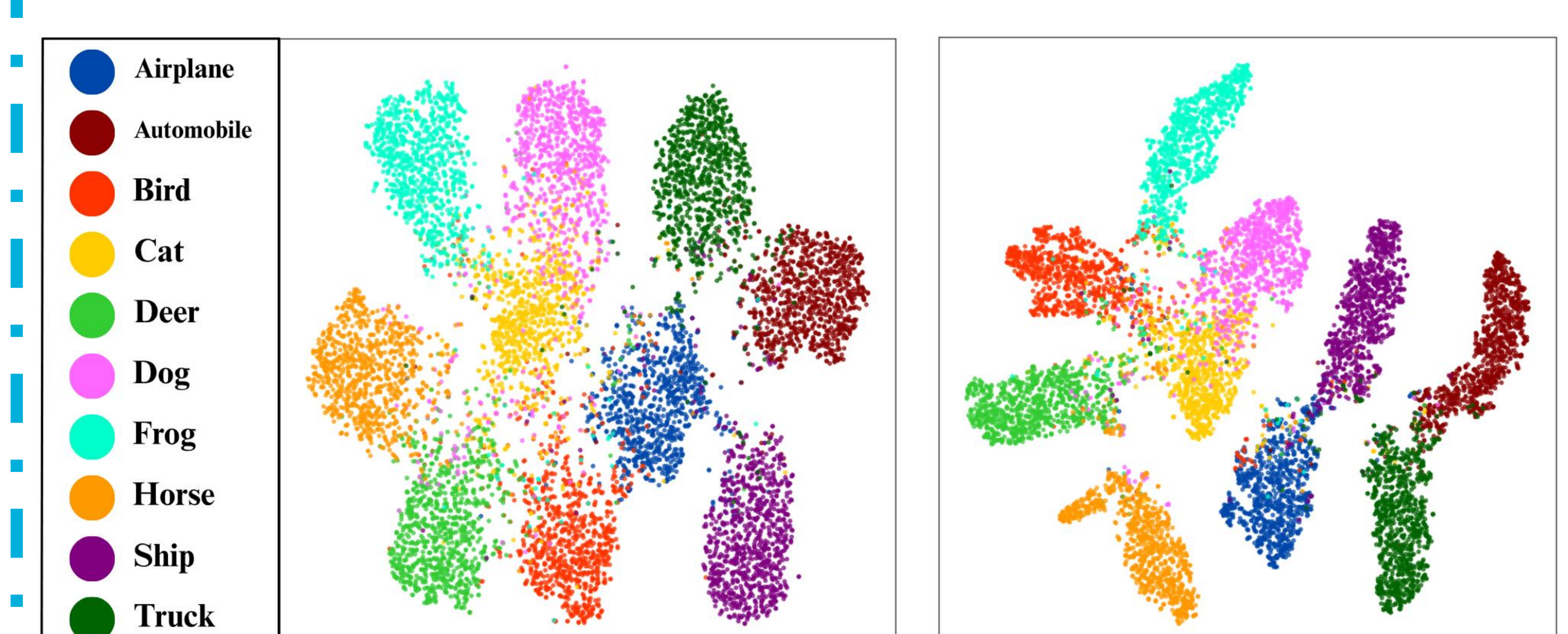
For a set of layers $\{1, \ldots, K\}$, compute a graph consistency loss at each layer and combine them:

$$\mathcal{L}_{\mathrm{GCR}} = \sum_{l=1}^{K} w_l \|\mathrm{triu}(F^{(l)}) - \mathrm{triu}(P)\|_F^2, \tag{6}$$

$$\mathcal{L}_{\mathrm{total}} = \mathcal{L}_{\mathrm{CE}} + \lambda\, \mathcal{L}_{\mathrm{GCR}}$$

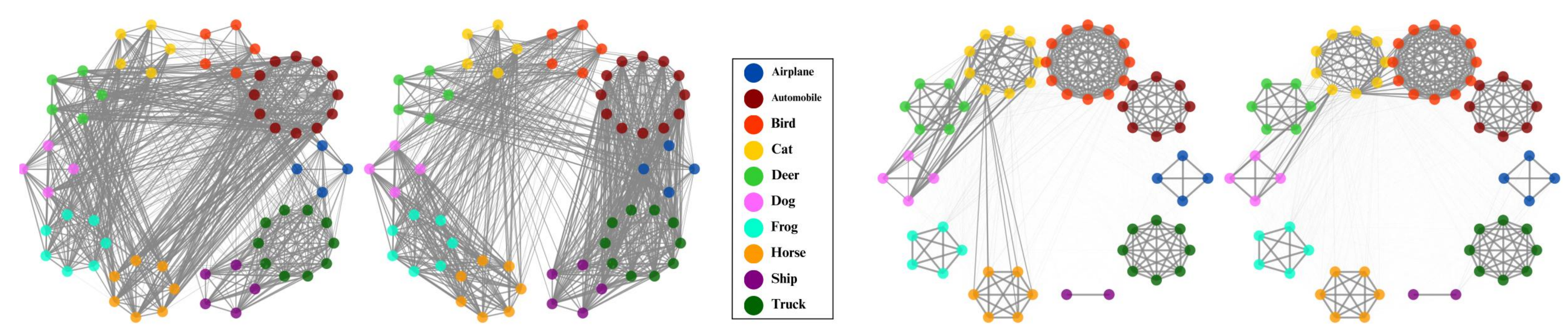
Results





	MAE	MNet	SN	SQNet	GLNet	Rx-50	Rx-101	R34	R50	R101	D121	Mean
Baseline	$88.95_{\pm0.33}$	$90.23_{\pm0.25}$	$91.21_{\pm0.28}$	$92.30_{\pm0.25}$	$94.10_{\pm0.26}$	$94.57_{\pm0.29}$	$95.12_{\pm0.30}$	$94.83_{\pm0.25}$	$95.03_{\pm0.28}$	$95.22_{\pm0.31}$	$95.01_{\pm0.27}$	$93.32_{\pm2.26}$
Early GCL	$89.42_{\pm0.25}$	$91.17_{\pm0.22}$	$92.33_{\pm0.33}$	$92.59_{\pm0.21}$	$\mathbf{94.89}_{\pm0.23}$	$95.48_{\pm0.22}$	$\underline{95.63}_{\pm0.25}$	$\underline{95.55}_{\pm0.18}$	$95.57_{\pm0.23}$	$95.39_{\pm0.26}$	$\mathbf{95.81}_{\pm0.17}$	$93.98_{\pm2.22}$
Mid GCL	$\mathbf{89.77}_{\pm0.22}$	$91.15_{\pm0.18}$	$\underline{92.58}_{\pm0.19}$	$92.40_{\pm0.20}$	$94.82_{\pm0.21}$	$95.47_{\pm0.19}$	$95.39_{\pm0.24}$	$\mathbf{95.69}_{\pm0.23}$	$95.61_{\pm0.20}$	$\mathbf{95.75}_{\pm0.17}$	$95.51_{\pm0.22}$	$\underline{94.01}_{\pm2.15}$
Late GCL	$\underline{89.70}_{\pm0.29}$	$\mathbf{91.40}_{\pm0.19}$	$92.36_{\pm0.21}$	$\underline{92.80}_{\pm0.19}$	$\underline{94.88}_{\pm0.19}$	$95.35_{\pm0.28}$	$\mathbf{95.71}_{\pm0.26}$	$\mathbf{95.69}_{\pm0.19}$	$\mathbf{95.66}_{\pm0.17}$	$\underline{95.51}_{\pm0.24}$	$\underline{95.72}_{\pm0.22}$	$\mathbf{94.07}_{\pm2.14}$
Early+Mid	$89.52_{\pm0.19}$	$90.77_{\pm0.26}$	$92.56_{\pm0.21}$	$92.27_{\pm0.25}$	$94.79_{\pm0.18}$	$95.33_{\pm0.27}$	$95.55_{\pm0.23}$	$95.46_{\pm0.20}$	$95.51_{\pm0.21}$	$95.37_{\pm0.19}$	$95.64_{\pm0.20}$	$93.89_{\pm2.22}$
Mid+Late	$89.59_{\pm0.28}$	$\underline{91.23}_{\pm0.20}$	$\mathbf{92.79}_{\pm0.20}$	$\mathbf{92.86}_{\pm0.23}$	$94.61_{\pm0.22}$	$\mathbf{95.51}_{\pm0.19}$	$95.38_{\pm0.27}$	$95.45_{\pm0.18}$	$95.33_{\pm0.26}$	$95.52_{\pm0.14}$	$95.70_{\pm0.19}$	$94.00_{\pm2.09}$
Early+Late	$89.64_{\pm0.25}$	$91.03_{\pm0.24}$	$92.30_{\pm0.28}$	$92.70_{\pm0.23}$	$94.69_{\pm0.20}$	$95.40_{\pm0.20}$	$95.35_{\pm0.23}$	$95.66_{\pm0.21}$	$95.31_{\pm0.25}$	$95.49_{\pm0.16}$	$95.53_{\pm0.22}$	$93.92_{\pm2.14}$
Full GCL	$89.55_{\pm0.23}$	$90.99_{\pm0.18}$	$92.48_{\pm0.19}$	$92.65_{\pm0.20}$	$94.57_{\pm0.21}$	$\underline{95.50}_{\pm0.19}$	$95.34_{\pm0.20}$	$95.48_{\pm0.17}$	$\underline{95.62}_{\pm0.18}$	$95.38_{\pm0.21}$	$95.51_{\pm0.20}$	$93.92_{\pm2.15}$

	MAE	MNet	SN	SQNet	Rx-50	Rx-101	R34	R50	D121	Mean
Baseline	$64.29_{\pm0.34}$	$65.95_{\pm0.25}$	$70.11_{\pm0.30}$	$69.43_{\pm0.27}$	$77.75_{\pm0.29}$	$77.83_{\pm0.30}$	$76.82_{\pm0.28}$	$77.31_{\pm0.29}$	$77.09_{\pm0.27}$	$72.95_{\pm5.50}$
Early GCL	$65.05_{\pm0.29}$	$67.45_{\pm0.21}$	$\mathbf{71.96}_{\pm0.27}$	$\underline{70.90}_{\pm0.20}$	$\underline{79.18}_{\pm0.22}$	$\underline{79.69}_{\pm0.27}$	$77.90_{\pm0.22}$	$\underline{79.37}_{\pm0.25}$	$79.41_{\pm0.22}$	$\underline{74.55}_{\pm5.78}$
Mid GCL	$64.99_{\pm0.30}$	$67.88_{\pm0.21}$	$\underline{71.89}_{\pm0.24}$	$70.21_{\pm0.25}$	$79.07_{\pm0.19}$	$79.28_{\pm0.26}$	$77.83_{\pm0.20}$	$78.90_{\pm0.24}$	$79.26_{\pm0.21}$	$74.37_{\pm5.66}$
Late GCL	$\mathbf{65.54}_{\pm0.27}$	$\underline{68.32}_{\pm0.20}$	$71.42_{\pm0.24}$	$70.55_{\pm0.22}$	$\mathbf{79.54}_{\pm0.20}$	$\mathbf{79.83}_{\pm0.21}$	$\mathbf{78.31}_{\pm0.20}$	$\mathbf{79.42}_{\pm0.21}$	$\mathbf{79.69}_{\pm0.23}$	$\mathbf{74.74}_{\pm5.73}$
Early+Mid	$65.23_{\pm0.31}$	$67.62_{\pm0.24}$	$71.50_{\pm0.28}$	$70.47_{\pm0.19}$	$78.90_{\pm0.18}$	$79.25_{\pm0.20}$	$77.41_{\pm0.19}$	$78.58_{\pm0.24}$	$79.22_{\pm0.20}$	$74.28_{\pm5.56}$
Mid+Late	$65.27_{\pm0.28}$	$\mathbf{68.33}_{\pm0.19}$	$71.63_{\pm0.28}$	$70.30_{\pm0.22}$	$78.91_{\pm0.17}$	$79.57_{\pm0.21}$	$77.30_{\pm0.20}$	$78.85_{\pm0.22}$	$\underline{79.54}_{\pm0.24}$	$74.41_{\pm5.55}$
Early+Late	$65.22_{\pm0.21}$	$67.25_{\pm0.21}$	$71.55_{\pm0.27}$	$\mathbf{71.03}_{\pm0.24}$	$79.03_{\pm0.20}$	$79.41_{\pm0.22}$	$\underline{78.19}_{\pm0.23}$	$78.70_{\pm0.23}$	$79.45_{\pm0.22}$	$74.43_{\pm5.69}$
Full GCL	$\underline{65.38}_{\pm0.22}$	$68.22_{\pm0.19}$	$71.30_{\pm0.24}$	$70.77_{\pm0.20}$	$79.01_{\pm0.19}$	$79.29_{\pm0.21}$	$77.79_{\pm0.20}$	$78.71_{\pm0.22}$	$79.27_{\pm0.19}$	$74.42_{\pm5.49}$



(a) DenseNet-121 (b) With our GCLs (c) MobileNet (d) With our GCLs

The relational graphs show that adding GCLs yields cleaner, tighter class clusters with fewer cross-class links, reducing feature noise and aligning features with semantic predictions