

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?

Method

We use cosine similarity with non-negative values:

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

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

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

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

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

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

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

$$P_{ij} = M_{ij} \odot S_{ij}, \quad (4)$$

where \odot denotes elementwise multiplication.

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

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

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

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

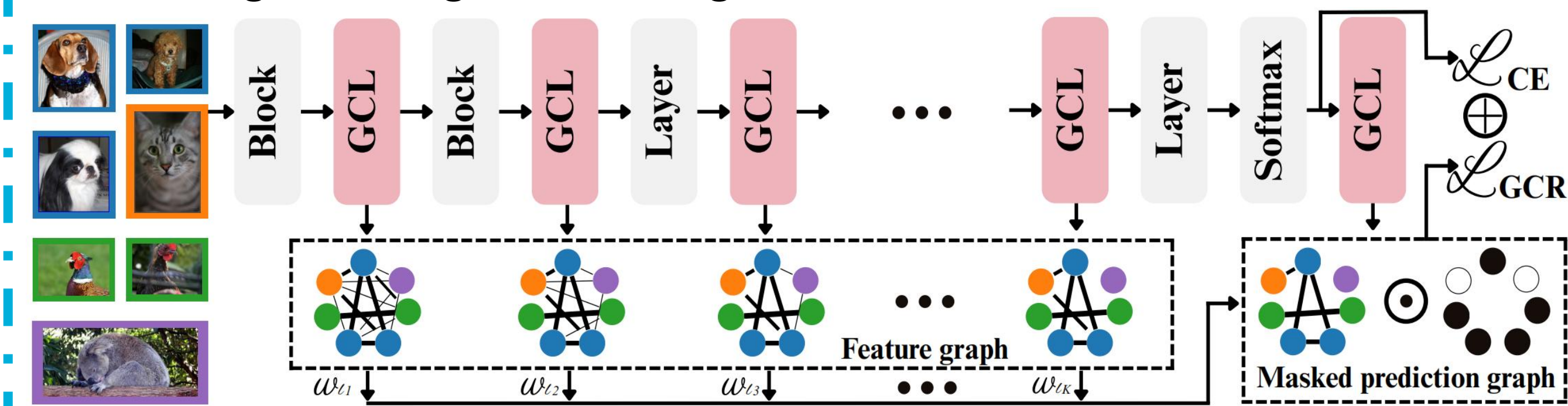
$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{CE}} + \lambda \mathcal{L}_{\text{GCR}}$$

Contributions

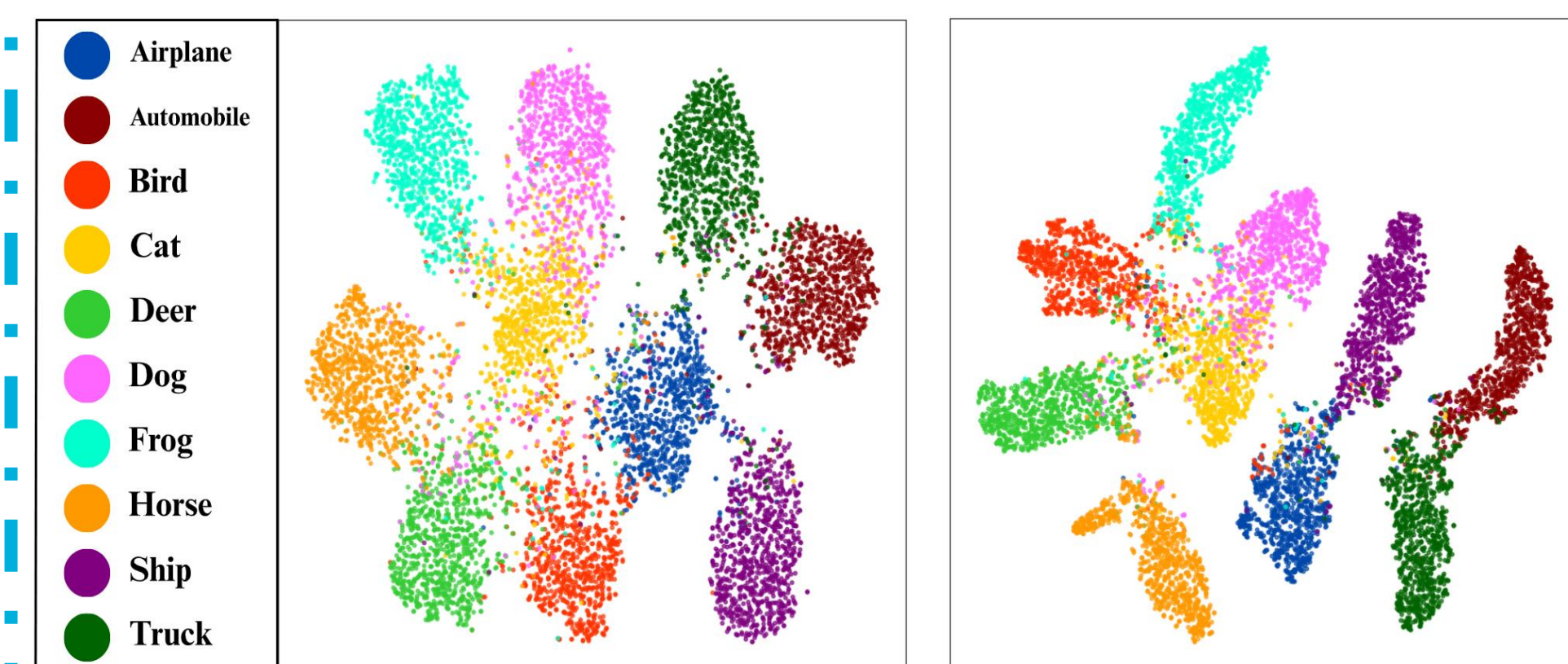
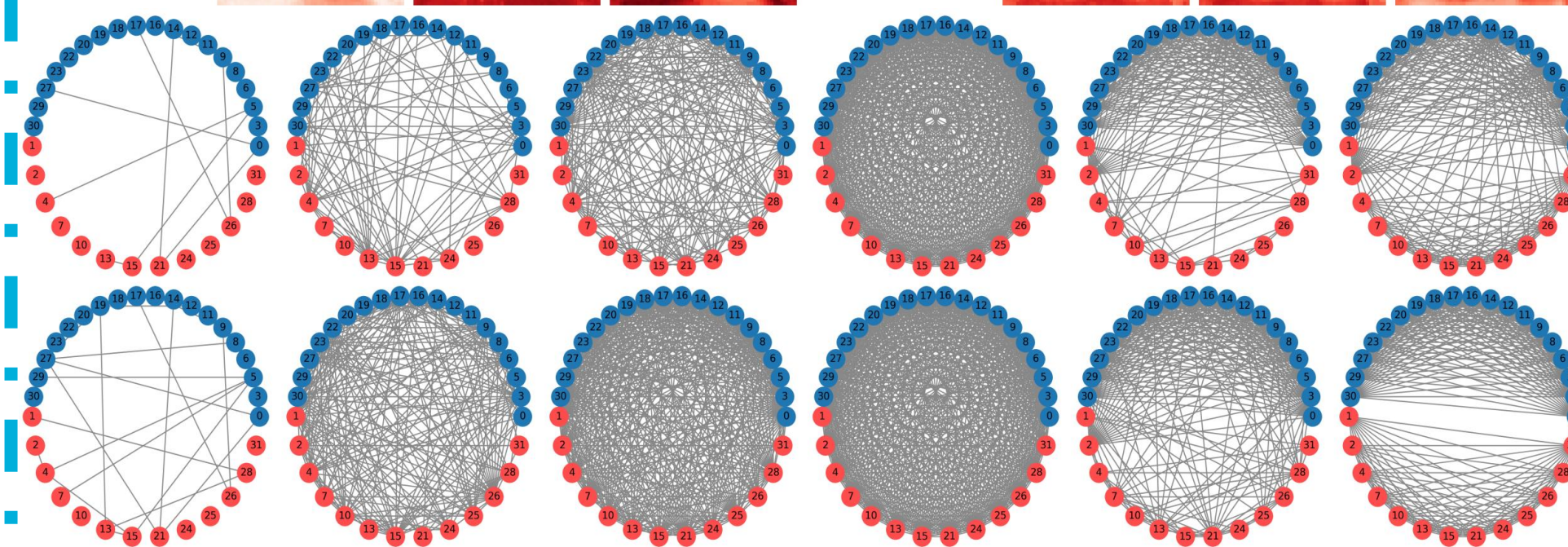
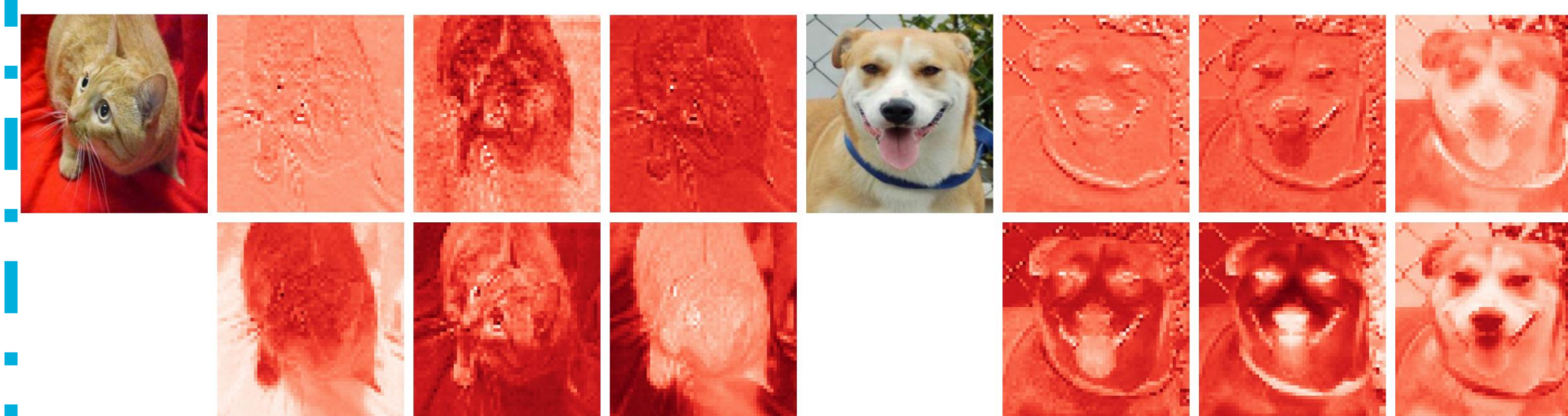
- GCR**: A new framework that aligns feature relational graphs with a class-aware masked prediction graph to enforce semantic consistency.
- GCLs**: Parameter-free layers that inject multi-layer relational supervision.
- Cross-space alignment**: Dynamically weights layers to refine semantic structure.
- Model-agnostic improvements**: Consistent boosts in semantic coherence and generalization across diverse architectures.

Pipeline

- Insert parameter-free GCLs after any block or layer.
- Each GCL builds a feature relational graph from batch features.
- Softmax outputs form a masked prediction graph, keeping only semantically consistent pairs.
- GCLs align feature graphs to prediction graphs.
- Alignment signals are weighted and added to CE loss.



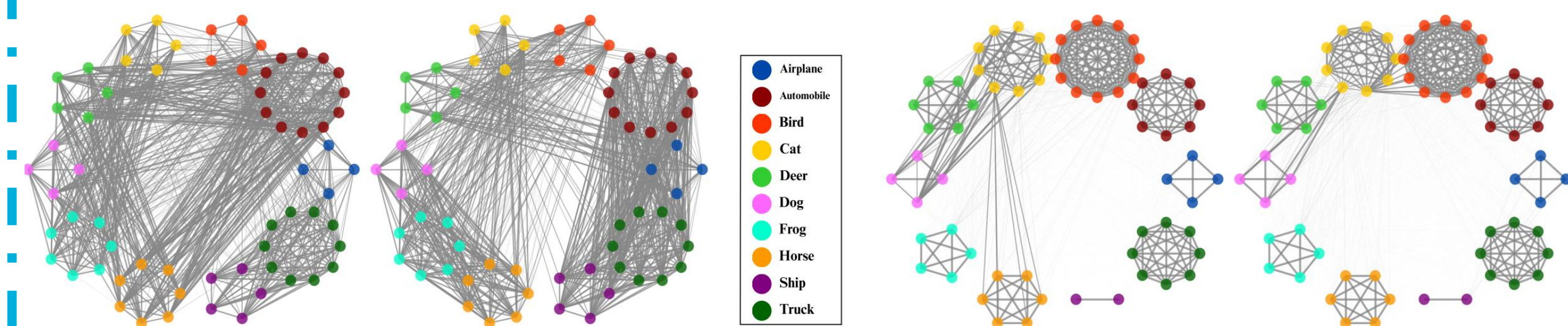
Results



Results

	MAE	MNet	SN	SQNet	GLNet	Rx-50	Rx-101	R34	R50	R101	D121	Mean
Baseline	88.95 \pm 0.33	90.23 \pm 0.25	91.21 \pm 0.28	92.30 \pm 0.25	94.10 \pm 0.26	94.57 \pm 0.29	95.12 \pm 0.30	94.83 \pm 0.25	95.03 \pm 0.28	95.22 \pm 0.31	95.01 \pm 0.27	93.32 \pm 2.26
Early GCL	89.42 \pm 0.25	91.17 \pm 0.22	92.33 \pm 0.33	92.59 \pm 0.21	94.89 \pm 0.23	95.48 \pm 0.22	95.63 \pm 0.25	95.55 \pm 0.18	95.57 \pm 0.23	95.39 \pm 0.26	95.81 \pm 0.17	93.98 \pm 2.22
Mid GCL	89.77 \pm 0.22	91.15 \pm 0.18	92.58 \pm 0.19	92.40 \pm 0.20	94.82 \pm 0.21	95.47 \pm 0.19	95.39 \pm 0.24	95.69 \pm 0.23	95.61 \pm 0.20	95.75 \pm 0.17	95.51 \pm 0.22	94.01 \pm 2.15
Late GCL	89.70 \pm 0.29	91.40 \pm 0.19	92.36 \pm 0.21	92.80 \pm 0.19	94.88 \pm 0.19	95.35 \pm 0.28	95.71 \pm 0.26	95.69 \pm 0.19	95.66 \pm 0.17	95.51 \pm 0.24	95.72 \pm 0.22	94.07 \pm 2.14
Early+Mid	89.52 \pm 0.19	90.77 \pm 0.26	92.56 \pm 0.21	92.27 \pm 0.25	94.79 \pm 0.18	95.33 \pm 0.27	95.55 \pm 0.23	95.46 \pm 0.20	95.51 \pm 0.21	95.37 \pm 0.19	95.64 \pm 0.20	93.89 \pm 2.22
Mid+Late	89.59 \pm 0.28	91.23 \pm 0.20	92.79 \pm 0.20	92.86 \pm 0.23	94.61 \pm 0.22	95.51 \pm 0.19	95.38 \pm 0.27	95.45 \pm 0.18	95.33 \pm 0.26	95.52 \pm 0.14	95.70 \pm 0.19	94.00 \pm 2.09
Early+Late	89.64 \pm 0.25	91.03 \pm 0.24	92.30 \pm 0.28	92.70 \pm 0.23	94.69 \pm 0.20	95.40 \pm 0.20	95.35 \pm 0.23	95.66 \pm 0.21	95.31 \pm 0.25	95.49 \pm 0.16	95.53 \pm 0.22	93.92 \pm 2.14
Full GCL	89.55 \pm 0.23	90.99 \pm 0.18	92.48 \pm 0.19	92.65 \pm 0.20	94.57 \pm 0.21	95.50 \pm 0.19	95.34 \pm 0.20	95.48 \pm 0.17	95.62 \pm 0.18	95.38 \pm 0.21	95.51 \pm 0.20	93.92 \pm 2.15

	MAE	MNet	SN	SQNet	Rx-50	Rx-101	R34	R50	D121	Mean
Baseline	64.29 \pm 0.34	65.95 \pm 0.25	70.11 \pm 0.30	69.43 \pm 0.27	77.75 \pm 0.29	77.83 \pm 0.30	76.82 \pm 0.28	77.31 \pm 0.29	77.09 \pm 0.27	72.95 \pm 5.50
Early GCL	65.05 \pm 0.29	67.45 \pm 0.21	71.96 \pm 0.27	70.90 \pm 0.20	79.18 \pm 0.22	79.69 \pm 0.27	77.90 \pm 0.22	79.37 \pm 0.25	79.41 \pm 0.22	74.55 \pm 5.78
Mid GCL	64.99 \pm 0.30	67.88 \pm 0.21	71.89 \pm 0.24	70.21 \pm 0.25	79.07 \pm 0.19	79.28 \pm 0.26	77.83 \pm 0.20	78.90 \pm 0.24	79.26 \pm 0.21	74.37 \pm 5.66
Late GCL	65.54 \pm 0.27	68.32 \pm 0.20	71.42 \pm 0.24	70.55 \pm 0.22	79.54 \pm 0.20	79.83 \pm 0.21	78.31 \pm 0.20	79.42 \pm 0.21	79.69 \pm 0.23	74.74 \pm 5.73
Early+Mid	65.23 \pm 0.31	67.62 \pm 0.24	71.50 \pm 0.28	70.47 \pm 0.19	78.90 \pm 0.18	79.25 \pm 0.20	77.41 \pm 0.19	78.58 \pm 0.24	79.22 \pm 0.20	74.28 \pm 5.56
Mid+Late	65.27 \pm 0.28	68.33 \pm 0.19	71.63 \pm 0.28	70.30 \pm 0.22	78.91 \pm 0.17	79.57 \pm 0.21	77.30 \pm 0.20	78.85 \pm 0.22	79.54 \pm 0.24	74.41 \pm 5.55
Early+Late	65.22 \pm 0.21	67.25 \pm 0.21	71.55 \pm 0.27	71.03 \pm 0.24	79.03 \pm 0.20	79.41 \pm 0.22	78.19 \pm 0.23	78.70 \pm 0.23	79.45 \pm 0.22	74.43 \pm 5.69
Full GCL	65.38 \pm 0.22	68.22 \pm 0.19	71.30 \pm 0.24	70.77 \pm 0.20	79.01 \pm 0.19	79.29 \pm 0.21	77.79 \pm 0.20	78.71 \pm 0.22	79.27 \pm 0.19	74.42 \pm 5.49



(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