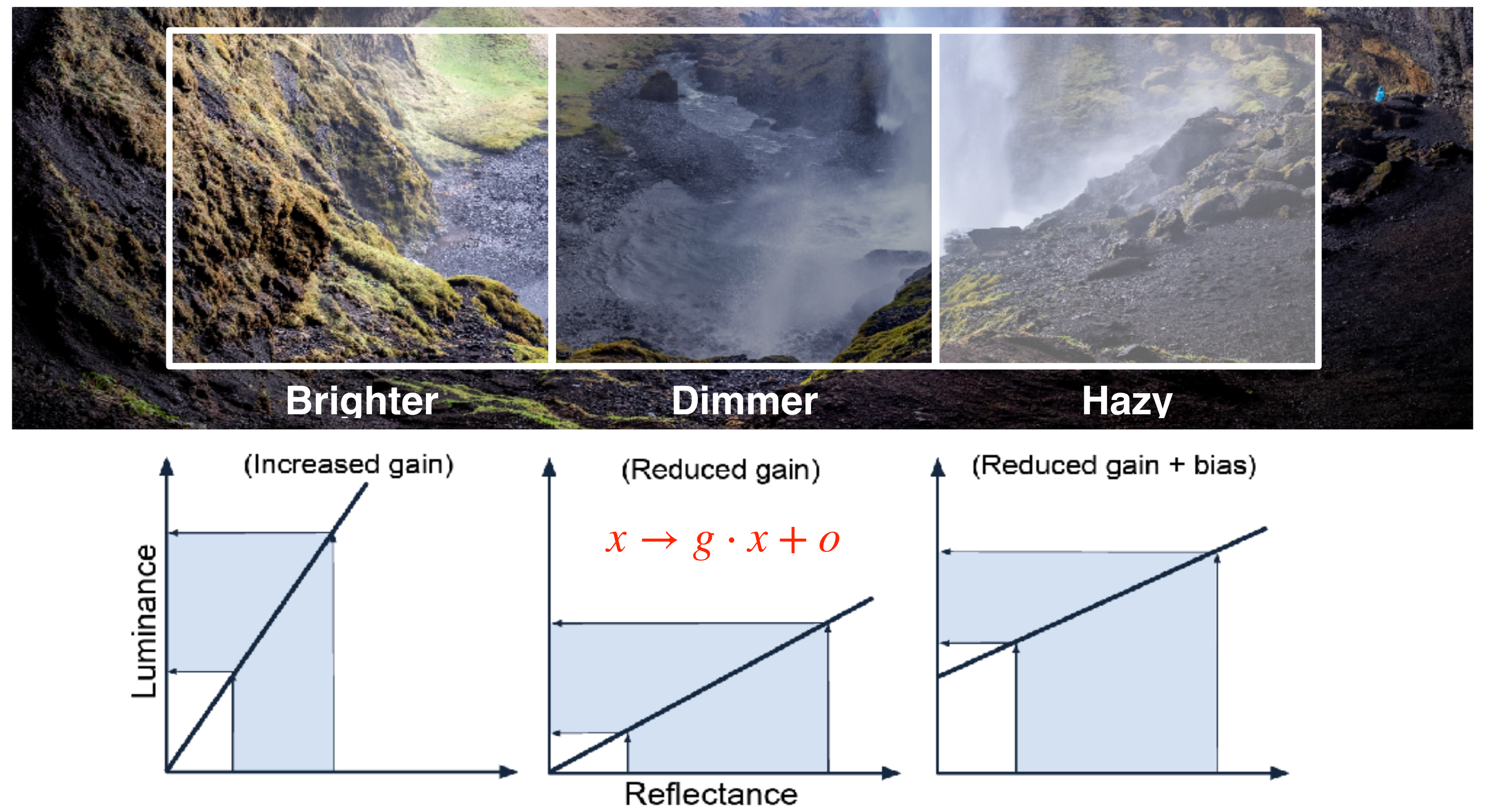
Normalize Filters! Classical Wisdom for Deep Vision

Gustavo Perez   Stella X. Yu

Atmosphere Changes Appearance, Not Semantics



Filter Normalization: Equivariance to Intensity Changes

$$f(x) = \sum w_i x_i \Rightarrow f(g \cdot x + o) = g \cdot f(x) + o \cdot \sum w_i \qquad \text{unnormalized}$$

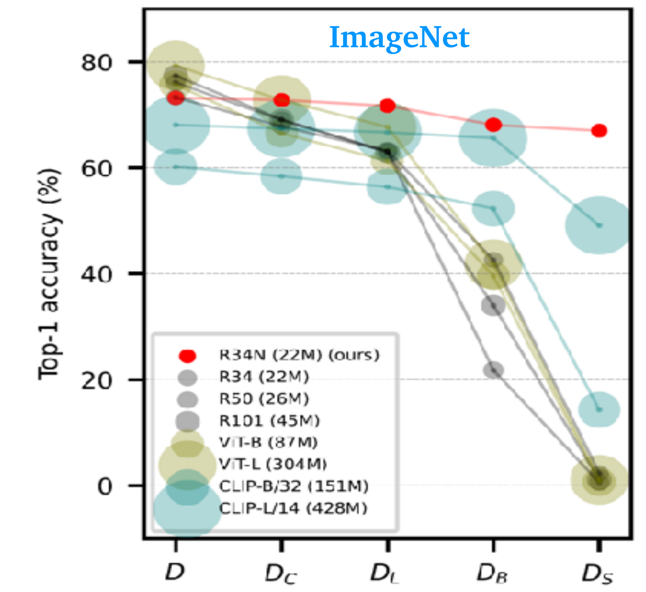
$$f(g \cdot x + o) = g \cdot f(x) + o, \quad \text{if } \sum w_i = 1 \qquad \text{averaging}$$

$$f(g \cdot x + o) = g \cdot f(x), \quad \text{if } \sum w_i = 0 \qquad \text{differencing}$$

$$y = \sum w_i x_i + b = \|w\| \sum \frac{w_i}{\|w\|} x_i + b = \|w\| \sum \left(\frac{w_i^+}{\|w\|} - \frac{w_i^-}{\|w\|} \right) x_i + b$$

$$\Rightarrow \underbrace{a}_{\text{scaling}} \underbrace{\sum \left(\frac{w_i^+}{\|w^+\| + \varepsilon} - \frac{w_i^-}{\|w^-\| + \varepsilon} \right)}_{\text{filter normalization}} x_i + \underbrace{b}_{\text{shifting}}$$

1. Gain from Filter Normalization, Not Corruption Training



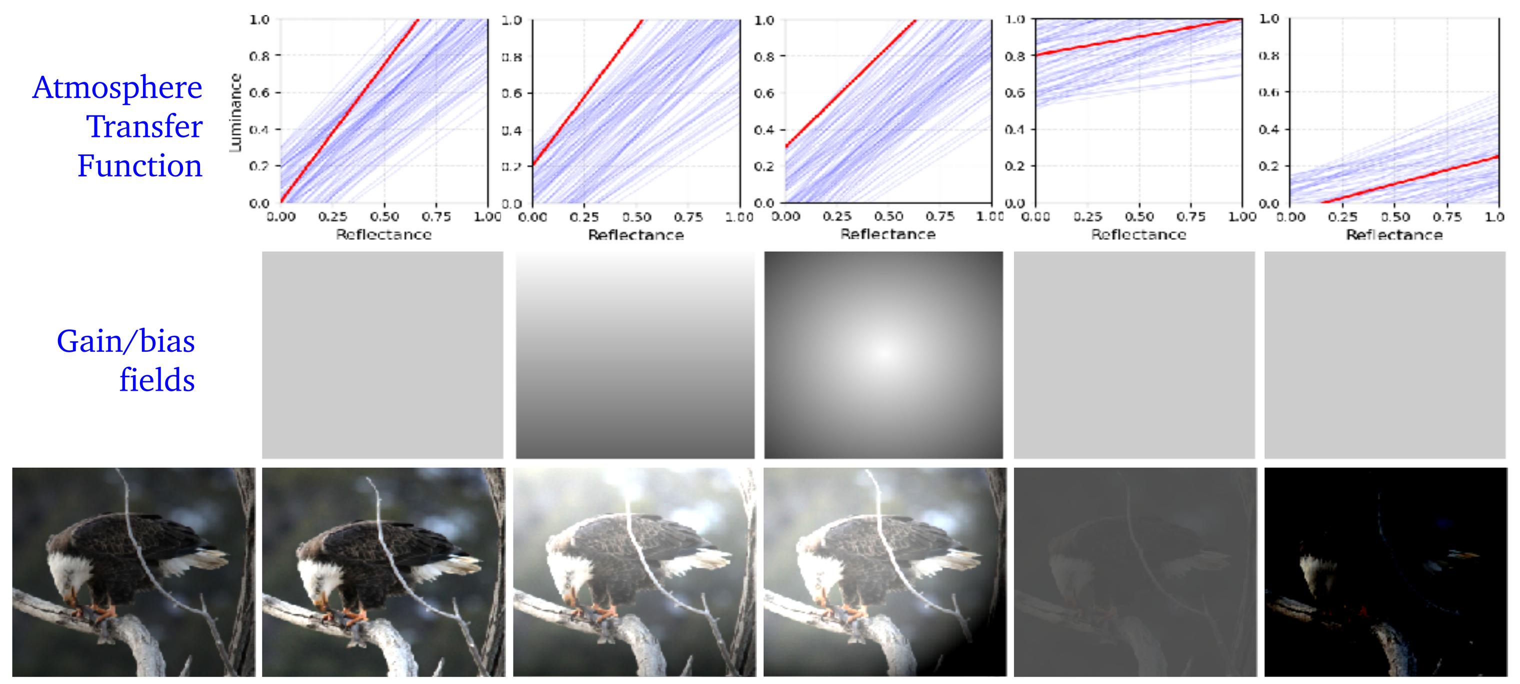
Model	# par	D	D_C	D_L	D_C	D_S
R34	22M	73.3	68.6	64.5	36.8	2.1
R50	26M	76.1	69.0	63.1	21.8	0.6
R101	45M	77.4	69.1	62.9	34.0	1.1
ViT-B	87M	75.7	66.5	61.3	39.6	0.7
ViT-L	304M	**79.3**	**72.8**	67.7	41.8	1.1
CLIP-B/32*	151M	60.2	58.4	56.4	52.3	14.3
CLIP-L/14*	428M	68.1	67.4	66.7	65.6	49.1
R34N	22M	73.2	**72.8**	**71.7**	**68.1**	**67.0**

*Trained on LAION-1B (zero-shot classification accuracy)

Train on D	D	D_C	D_L	D_C	D_S
CIFAR: **R20**	91.4	87.9	84.2	82.9	38.1
CIFAR: **R20N**	**91.5**	**91.1**	**89.6**	**85.5**	**89.5**

Train on D_C	D	D_C	D_L	D_C	D_S
CIFAR: **R20**	**91.9**	91.6	**90.8**	**88.0**	76.9
CIFAR: **R20N**	**91.9**	**91.7**	**90.8**	87.1	**90.8**

Reduces Recognition Even for Foundation Models



R34	✔	✔	✔	candle	torch	spotlight
ViT	✔	bald eagle	✔	nail	gray owl	hooks
CLIP	✔	✔	✔	✔	✔	bars
R34N	✔	✔	✔	✔	✔	✔

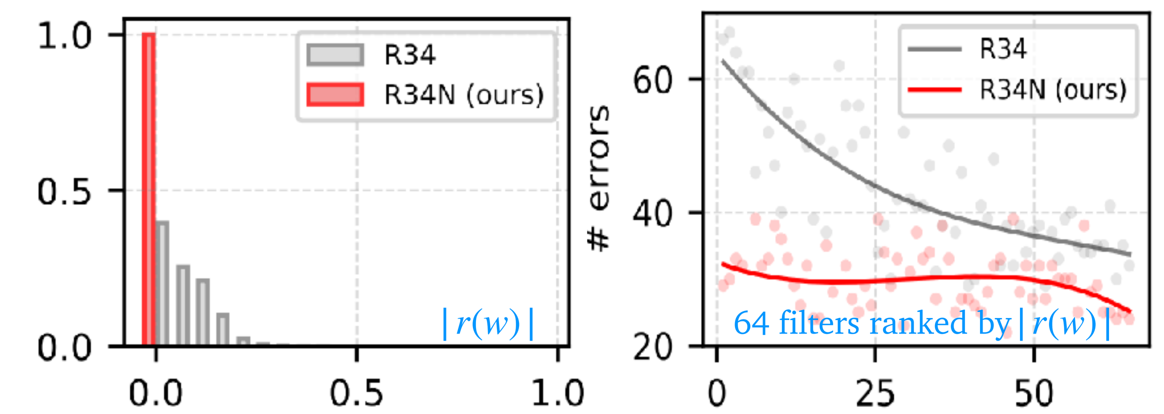
Normalized Filters: Fewer Errors, Greater Diversity

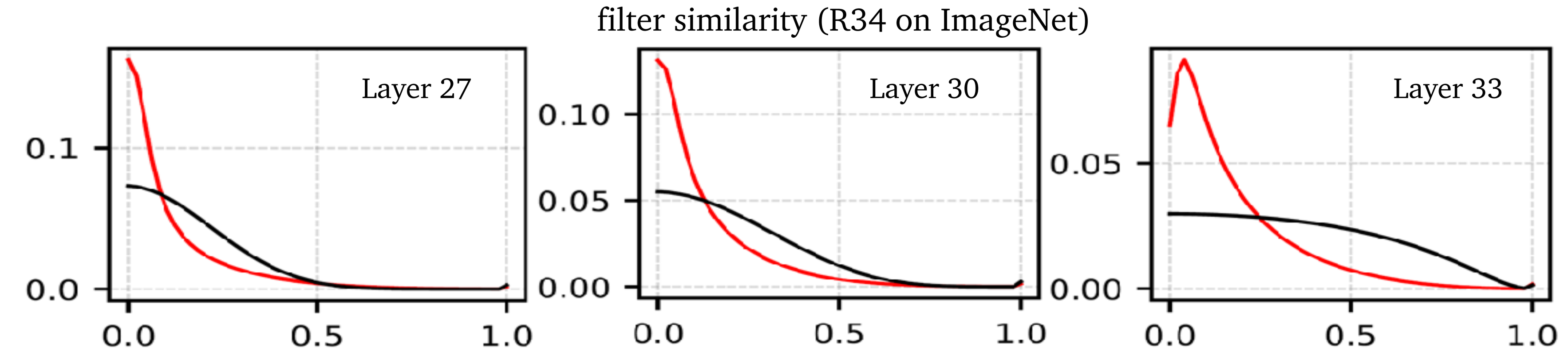
$$r(w) = \frac{\sum w_i}{\sum |w_i|}$$

$|r(w)| \le 1$ unnormalized

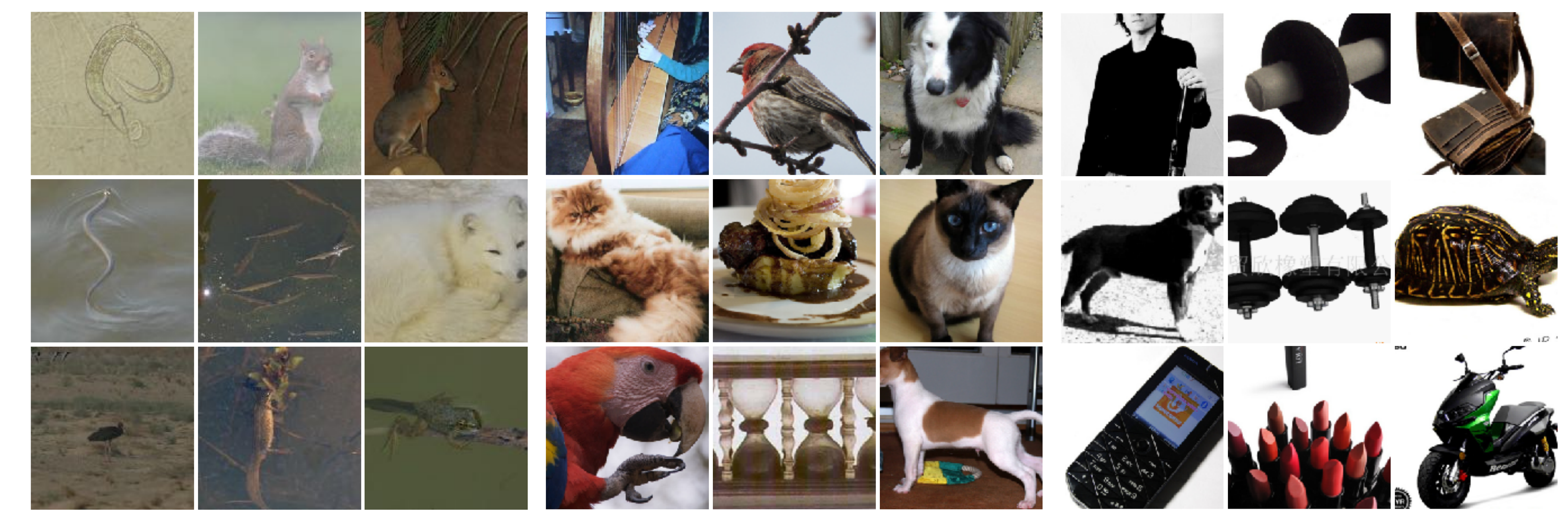
$r(w) = 1$ averaging

$r(w) = 0$ differencing



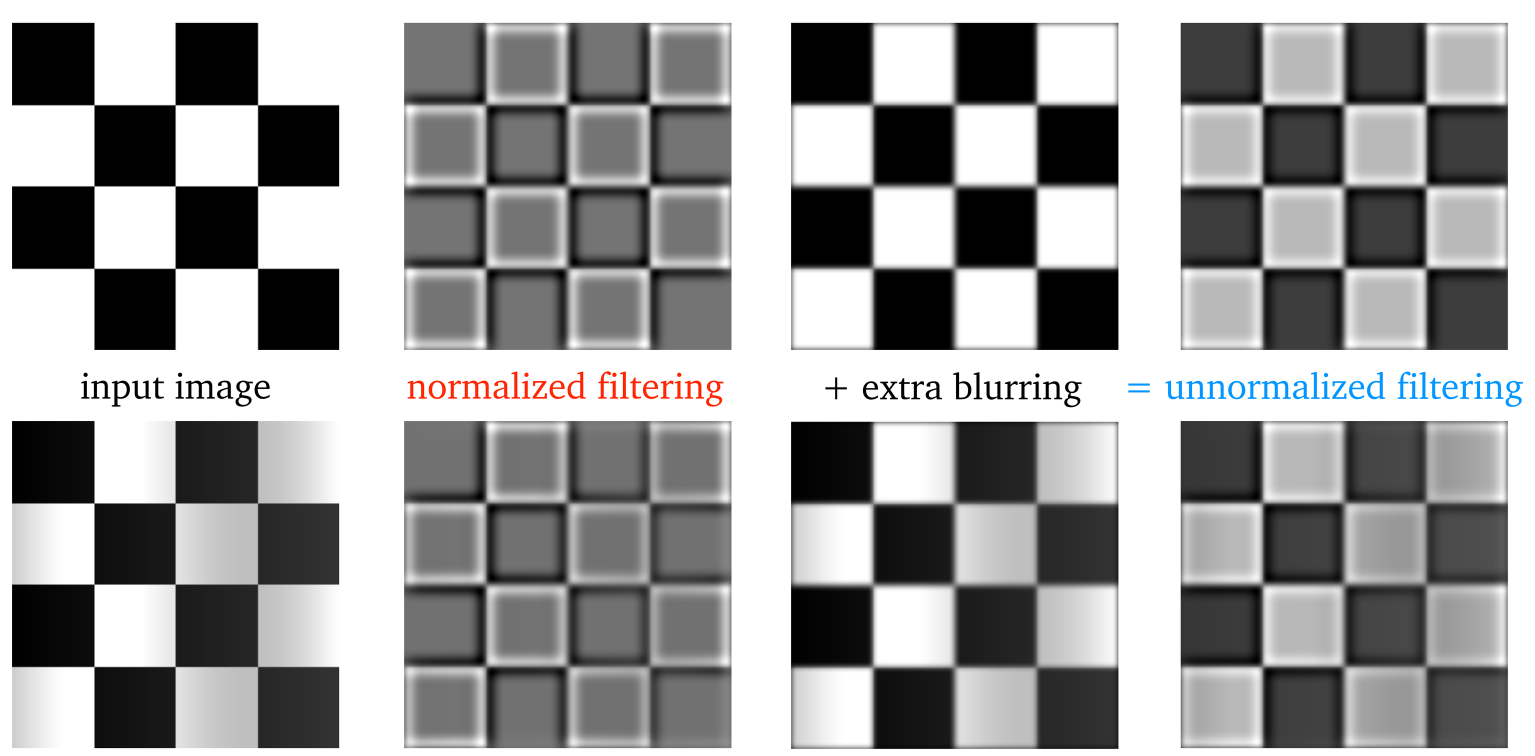


2. Robustness to Low/High-Constrast ImageNet Images

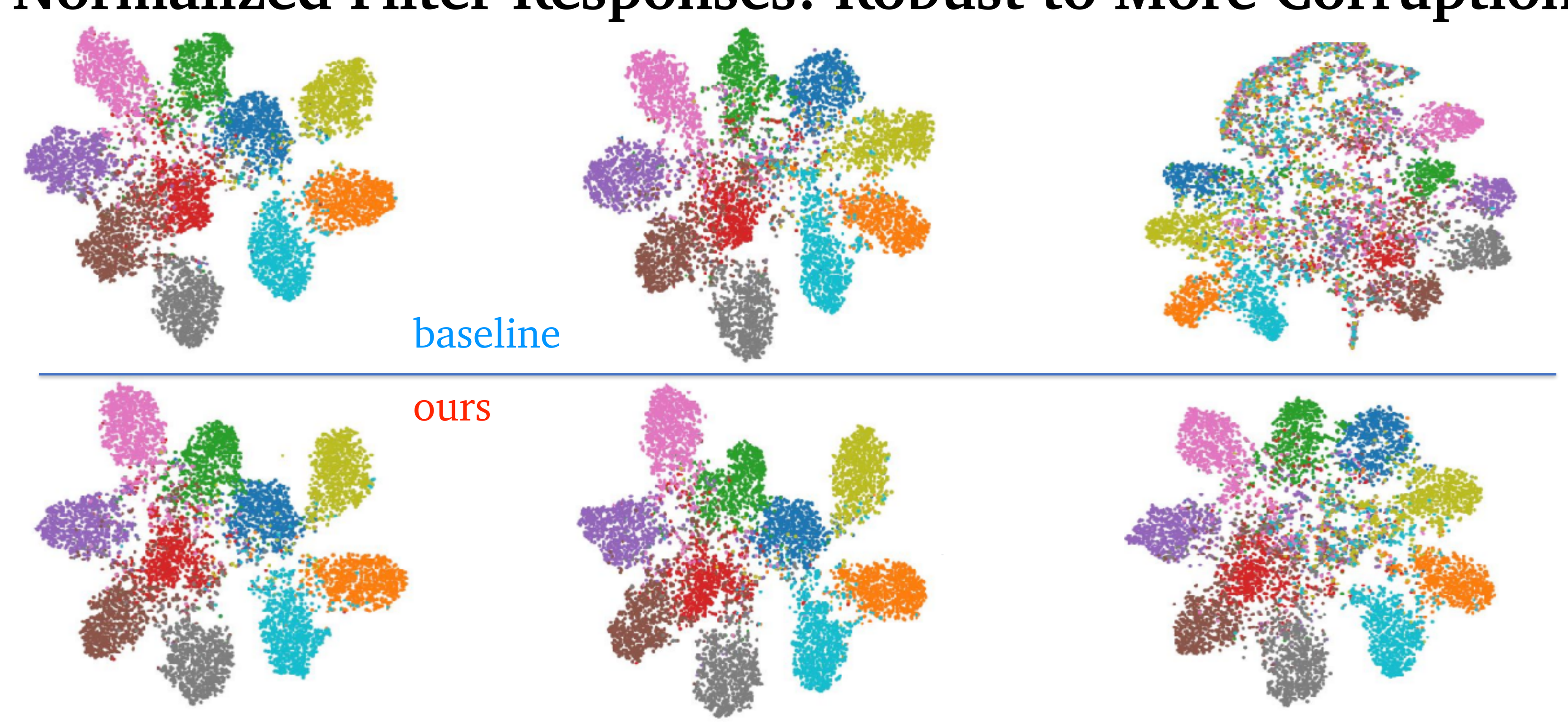


Model	Lowest				Contrast				Highest
R34	65.4	74.0	74.8	75.2	73.2	70.8	70.7	67.9	63.3
R34N	**69.2**	74.1	74.6	74.5	73.3	71.1	71.1	67.3	**71.4**

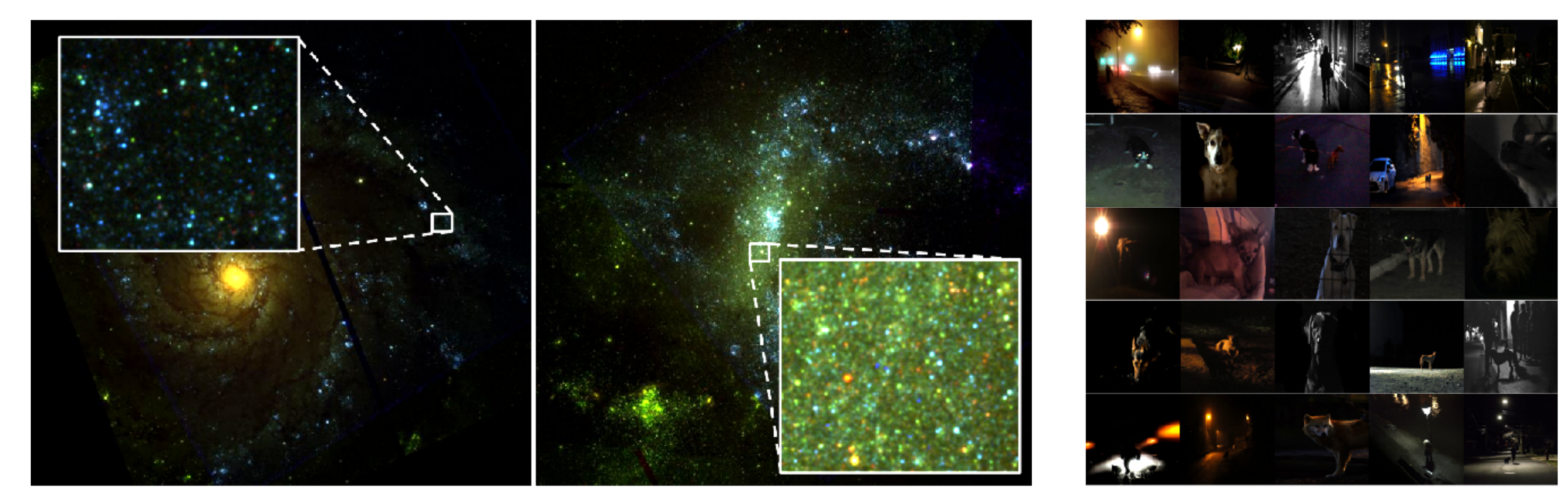
Normalize Filters for Consistency, Interpretability



Normalized Filter Responses: Robust to More Corruption



3. Better Generalization on Natural Extremely Dark Images



Dataset	Classification	Architecture	Base Accuracy	Our Accuracy (%)	Gain
LEGUS	star clusters	R18	50.2	51.9	+1.7
ExDark	objects	R34	28.3	34.2	+5.9