Scale-Equivariant U-Net: Supplementary Material

Mateus Sangalli
mateus.sangalli@minesparis.psl.eu
Samy Blusseau
samy.blusseau@minesparis.psl.eu
Santiago Velasco-Forero
santiago.velasco@minesparis.psl.eu
Jesús Angulo
jesus.angulo@minesparis.psl.eu

Mines Paris, PSL University
Centre for Mathematical Morphology
Fontainebleau, France

A Semigroup cross-correlation

Let $\mathcal{F} = \mathbb{R}^G$ denote the set of functions mapping G to $\mathbb{R}$. Bearing in mind the final purpose of defining equivariant CNN layers, we focus on linear operators on $\mathcal{F}$. Let the semigroup right action $(R_u)_{u\in G}$ on $\mathcal{F}$, defined by

$$\forall u, g \in G, \forall f \in \mathcal{F}, \quad R_u(f)(g) = f(u \cdot g). \tag{1}$$

Then for any $h \in \mathcal{F}$, the linear operator defined by

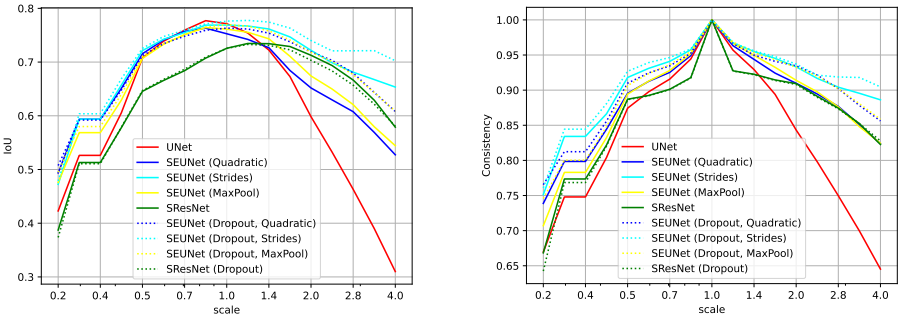
$$\forall u \in G, \quad H(f)(u) = (f \star_G h) = \sum_{g\in G} R_u(f)(g)h(g) \tag{2}$$

is equivariant to the semigroup action $(R_u)_{u\in G}$, as $H(R_u(f)) = R_u(H(f))$. This class of semigroup equivariant linear operators is the *semigroup cross-correlation* proposed in [3] as the key element to define scale-equivariant convolutional layers. Note that when G is the group of image translations (groups are special semigroups), (2) corresponds to the classic discrete convolution with the reversed filter $h^*(g) = h(g^{-1})$. We use the notation $f \star_G h$ remarking however that this operation is not symmetrical in f and h even when the law $\cdot$ on G is commutative. Also, contrary to the group case, we do not have the property that every linear and equivariant operator can be written as a semigroup-cross-correlation.

B Different Pooling Operators

Besides the strided scale-crosscorrelations we used, we can define another class of pooling operators, inspired by classical max-pooling. Let us place ourselves in a slightly different case of pooling a function in a continuous domain $f : \mathcal{S} \times \mathbb{R}^2 \to \mathbb{R}$, with $\mathcal{S} \times \mathbb{R}^2$ acting on it

© 2022. The copyright of this document resides with its authors.
It may be distributed unchanged freely in print or electronic forms.



(a) IoU (b) Consistency

Figure 1: Overall results in terms of IoU and Consistency for each scale of the Pet dataset. The SEU-Net has been computed with different pooling functions.

by $R_{\gamma^l,z}f(\gamma^k,x) = f(\gamma^{k+l},\gamma^l x+z)$, $k,l \in \mathbb{N}$, $x \in \mathbb{R}^2$, $z \in \mathbb{Z}^2$. We define the pooling of f as an operator F followed by a downsampling $D_{\gamma^l}[f](\gamma^k,z) = f(\gamma^k,\gamma^l z)$

$$P[f] = D_{\gamma^l} F f. \tag{3}$$

If F commutes with $R_{\gamma^k,x}$, then so does P. We consider three pooling functions: $F_{\text{id}} = \text{id}$ (strides) and two dilation scale-spaces [1]:

- The max-pooling of scale-semigroup-valued images is given by a re-scaled max-pooling

$$F_{\max}[f](\gamma^k,z) = \sup_{y \in N_k \times N_k} \{f(z-y)\} \tag{4}$$

  where $N_k = \{\gamma^k x | x \in N\}$ and N is for example a γ^l-sided square in $\mathbb{R}^2$.

- The quadratic dilation (quadpool) scale-space is a morphological counterpart to the Gaussian scale-space [2] defined by

$$F_{\text{quad}}[f](\gamma^k,z) = \sup_{y \in \mathbb{R}^2} \left\{ f(z-y) - \frac{\|y\|^2}{c\gamma^{2k}} \right\}, \tag{5}$$

  where $c > 0$ is some constant.

In contrast to the strided scale-cross-correlations given by F_{id}, the functions $F_{\max}$ and F_{quad} are scale-equivariant only in this continuous setting, their discretized versions are not actually equivariant. Nonetheless, a network employing scale-cross-correlations and these poolings would be equivariant when applied to signals in the domain $\mathcal{S} \times \mathbb{R}^2$.

In Figure 1 we extend the experiments from Section 5.2. Using different pooling functions did not improve the performance of the SEU-Net compared to its performance using strided scale-cross-correlations.

C Non-equivariance of the upsampling

In this section we show that $U_{\gamma^l} R_{\gamma^k,x} f \neq R_{\gamma^k,\gamma^l x} U_{\gamma^l} f$ for at least one lifted image f, one couple of integers (k,l) and a point $x \in \mathbb{Z}^2$. Note that U_{γ^l} is an upsampling defined in the associated paper.

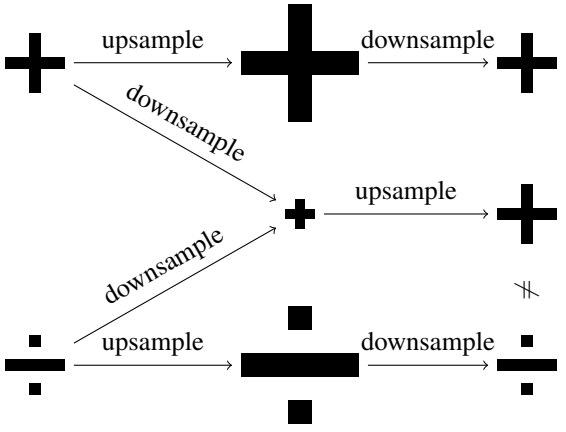


Figure 2: Example illustrating the problem with upsampling in a scale-equivariant architecture. We have images f_1 and f_2 such that when both are downsampled and then upsampled they yield the same result, but if both are upsampled and then downsampled they yield different results.

Given any $k \in \mathbb{N}$, take $l = k$, $x = (0,0)$ and any two lifted images f_1 and f_2 that coincide on certain points,

$$f_1(s, \gamma^k y) = f_2(s, \gamma^k y) \quad \forall s \in \mathcal{S}, y \in \mathbb{Z}^2,$$

and are different elsewhere, as illustrated in Figure 2. Let us show that $U_{\gamma^l} R_{\gamma^k,x} f_i \neq R_{\gamma^k,\gamma^l x} U_{\gamma^l} f_i$ either for $i = 1$ or $i = 2$ or both. The set of points where f_1 and f_2 coincide implies in particular that $R_{\gamma^k,0} f_1 = R_{\gamma^k,0} f_2$. Then we have

$$R_{\gamma^k,0} U_{\gamma^k} f_1 \neq R_{\gamma^k,0} U_{\gamma^k} f_2,$$

as $R_{\gamma^k,0} U_{\gamma^k} f_i : (s,y) \mapsto f_i(\gamma^k s, y)$, and $f_1(\gamma^k s, y) \neq f_2(\gamma^k s, y)$ for $y \notin k\mathbb{Z}^2$. Note that $R_{\gamma^k,0} U_{\gamma^k}$ is nothing else than an upsampling followed by a downsampling, as in Figure 2.
Since, on the other hand, $R_{\gamma^k,0} f_1 = R_{\gamma^k,0} f_2$, we get

$$U_{\gamma^k} R_{\gamma^k,0} f_1 = U_{\gamma^k} R_{\gamma^k,0} f_2.$$

Hence, either $R_{k,0} U_k f_1 \neq U_k R_{k,0} f_1$ or $R_{k,0} U_k f_2 \neq U_k R_{k,0} f_2$ or both, proving our point.

D Proof of Proposition 1

Proof. First, consider $k < m$

$$\begin{aligned} R_{\gamma^k,0} \circ U_{\gamma^k} f(\gamma^p, y) &= (U_{\gamma^k} f)(\gamma^k \gamma^p, \gamma^k y) \\ &= f(\gamma^k \gamma^p, y) \end{aligned}$$

so $R_{\gamma^k,0} U_{\gamma^m} f(\gamma^p, y) = U_{\gamma^{m-k}} f(\gamma^{p+m}, y)$.

Now, let $f = U_{\gamma^m} f_i \in F$, $k \leq \min\{n_i | i = 1, \ldots, N\} \leq m$, we have

$$\begin{aligned} U_\gamma R_{\gamma^k,x} f(\gamma^p, y) &= U_\gamma R_{\gamma^k,x} U_{\gamma^m} f_i(\gamma^p, y) \\ &= U_\gamma R_{1,x} R_{\gamma^k,0} U_{\gamma^m} f_i(\gamma^p, y) \\ &= R_{1,\gamma x} U_\gamma U_{\gamma^{m-k}} f_i(\gamma^{p+m} r, y) \\ &= U_{\gamma^{m-k+1}} f_i(\gamma^{m+p}, y + \gamma x) \end{aligned}$$

and, on the other hand

$$\begin{aligned} R_{\gamma^k,\gamma x} U_\gamma f(\gamma^p, y) &= R_{\gamma^k,\gamma x} U_\gamma U_{\gamma^m} f_i(\gamma^p, y) \\ &= R_{1,\gamma x} R_{\gamma^k,0} U_{\gamma^{m+1}} f_i(\gamma^p, y) \\ &= R_{1,\gamma x} U_{\gamma^{m-k+1}} f_i(\gamma^{p+k}, y) \\ &= U_{\gamma^{m-k+1}} f_i(\gamma^{p+m}, y + \gamma x) \\ &= U_\gamma R_{\gamma^k,x} f(\gamma^p, y), \end{aligned}$$

implying $U_\gamma R_{\gamma^k,x} f = R_{\gamma^k,\gamma x} U_\gamma f$. Repeated application gives us the desired result. □

E Details of the Experiments

E.1 Oxford IIIT Pet

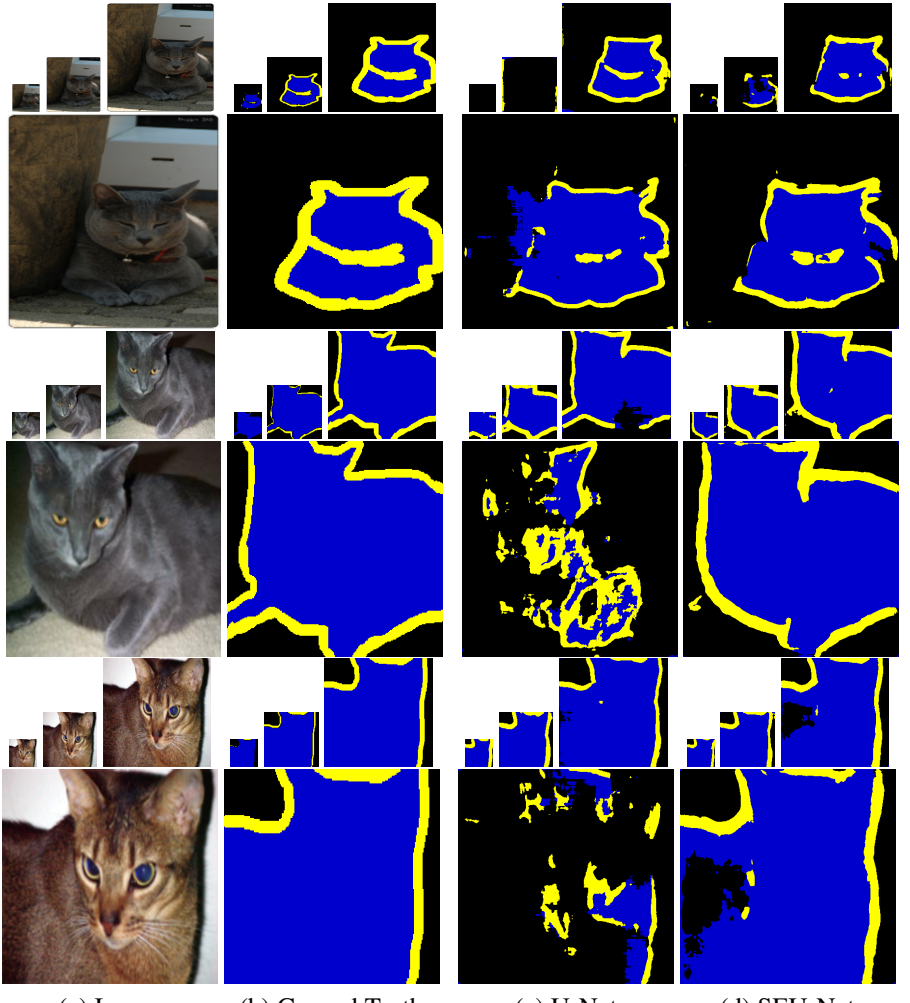
All models, except for the augmented U-Net are trained for 300 epochs. The augmented U-Net is trained for four times as many epochs. To train all models we apply data augmentation consisting of, rotations by a uniformly sampled angles in $[-10°, 10°]$, linear contrast changes by values in the range $[0.9, 1.1]$, random horizontal flipping and random cropping to size 112×112. Learning rate starts at 10^{-3} an is reduced by 10 when the validation loss does not improve for 30 epochs. We use a batch size of 8.

E.2 DIC-C2DH-HeLa

All models, except for the U-Net with jittering 4 are trained for 200 epochs. The augmented U-Net with jittering 4 is trained for four times as many epochs. To train all models we apply data augmentation consisting of, rotations by a uniformly sampled angles in $[-10°, 10°]$, linear contrast changes by values in the range $[0.9, 1.1]$, random horizontal and vertical flipping and elastic transformations. Learning rate starts at 10^{-3}, weight decay starts at 10^{-4} and both are reduced by exponential decay such that they are divided by 10 every 100 epochs (the decay stops at epoch 300 for the U-Net with size 4 jittering). We use a batch size of 1.

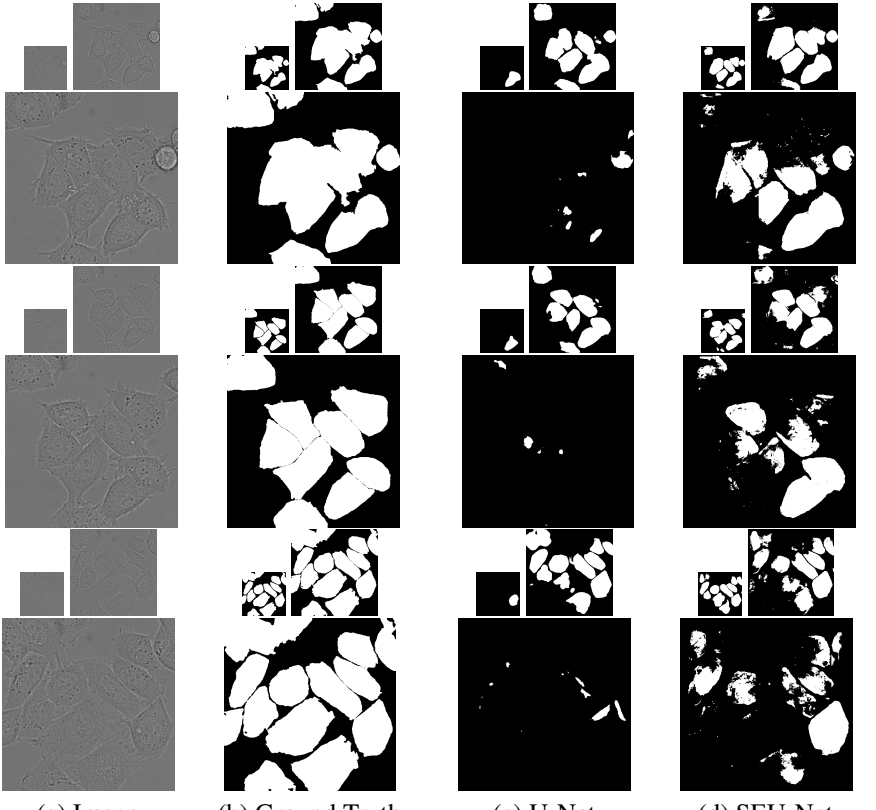
F More Examples of Predictions

In Figure 3 we can see some more examples of predictions from the Oxford Pet dataset, particularly when the U-Net struggles to generalize to new scales. Similarly Figure 4 showcases some extra examples from the experiment from the DIC-HeLa experiment.



(a) Image (b) Ground Truth (c) U-Net (d) SEU-Net

Figure 3: Sample test images at different scales and ground truth from the Oxford-IIIT Pet dataset, along with the U-Net and SEU-Net predictions. The scales present are $0.25, 0.5, 1$ and 2 times the training scale.



(a) Image (b) Ground Truth (c) U-Net (d) SEU-Net

Figure 4: Predictions from DIC-HeLa at different scales, namely scales 0.5, 1 and 2.

References

[1] Henk JAM Heijmans and Rein van den Boomgaard. Algebraic framework for linear and morphological scale-spaces. *Journal of Visual Communication and Image Representation*, 13(1-2):269–301, 2002.

[2] Rein Van Den Boomgaard and Arnold Smeulders. The morphological structure of images: The differential equations of morphological scale-space. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(11):1101–1113, 1994.

[3] Daniel Worrall and Max Welling. Deep scale-spaces: Equivariance over scale. In *Advances in Neural Information Processing Systems*, pages 7364–7376, 2019.