How to Transform Kernels for Scale-Convolutions

Anonymous ICCV submission

Paper ID ****

Abstract

Scale is often seen as a given, disturbing factor in many vision tasks. When doing so it is one of the factors why we need more data during learning. In recent work scale equivariance was added to convolutional neural networks. It was shown to be effective for a range of tasks. We aim for accurate scale-equivariant convolutional neural networks (SE-CNNs) applicable for problems where high granularity of scale and small kernel sizes are required. Current SE-CNNs rely on weight sharing and kernel rescaling, the latter of which is accurate for integer scales only. To reach accurate scale equivariance, we derive general constraints under which scale-convolution remains equivariant to discrete rescaling. We find the exact solution for all cases where it exists, and compute the approximation for the rest. The discrete scale-convolution pays off, as demonstrated in a new state-of-the-art classification on MNIST-scale and on STL-10 in the supervised learning setting.

1. Introduction

Scale is a natural attribute of every object, as basic property as location and appearance. And hence it is a factor in almost every task in computer vision. In image classification, global scale invariance plays an important role in achieving accurate results [16]. In image segmentation and object tracking, scale equivariance is important as the output map should scale proportionally to the input [1, 27]. Where scale invariance or equivariance is usually left as a property to learn in the training of these computer vision methods by providing a good variety of examples [20], we aim for accurate scale analysis for the purpose of needing less data to learn from.

Scale of the object can be derived externally from the size of its silhouette, e.g [36], or internally from the scale of its details, e.g [4]. External scale estimation requires the full object to be visible. It will easily fail when the object is occluded and/or when the object is amidst a cluttered background, for example for people in a crowd [26], when proper detection is hard. In contrast, internal scale estimation is build on the scale of common details [25], for example deriving the scale of a person from the scale of a sweater or a face. Where internal scale has better chances of being reliable, it poses heavier demands on the accuracy of assessment than external scale estimation. We focus on improvement of the accuracy of internal scale analysis.

We focus on accurate scale analysis on the generally applicable scale-equivariant convolutional neural networks [34, 3, 28]. A scale-equivariant network extends the equivariant property of conventional convolutions to the scale-translation group. It is achieved by rescaling the kernel basis and sharing weights between scales. While the weight sharing is defined by the structure of the group [9], the proper way to rescale kernels is an open problem. In [3, 28], the authors propose to rescale kernels in the continuous domain to project them later on a pixel grid. This permits the use of arbitrary scales, which is important to many application problems, but the procedure may cause a significant equivariance error [28]. Therefore, Worrall and Welling [34] model rescaling as a dilation, which guarantees a low equivariance error at the expense of permitting only integer scale factors. Due to the continuous nature of observed scale, integer scale factors may not cover the range of variations in the best possible way.

In the paper, we show how the equivariance error affects the performance of SE-CNNs. We make the following contributions:

- From first principles we derive the best kernels, which minimize the equivariance error.
- We find the conditions when the solution exists and find a good approximation when it does not exist.
- We demonstrate that an SE-CNN with the proposed kernels outperforms recent SE-CNNs in classification in both accuracy and compute time. We set new state-of-the-art results on MNIST-scale and STL-10.

The proposed approach contains [34] as a special case. Moreover, the proposed kernels can't be derived from [28] and vice versa. The union of our approach and the approach presented in [28] covers the whole set of possible SE-CNNs for a finite set of scales.

ICCV
#****

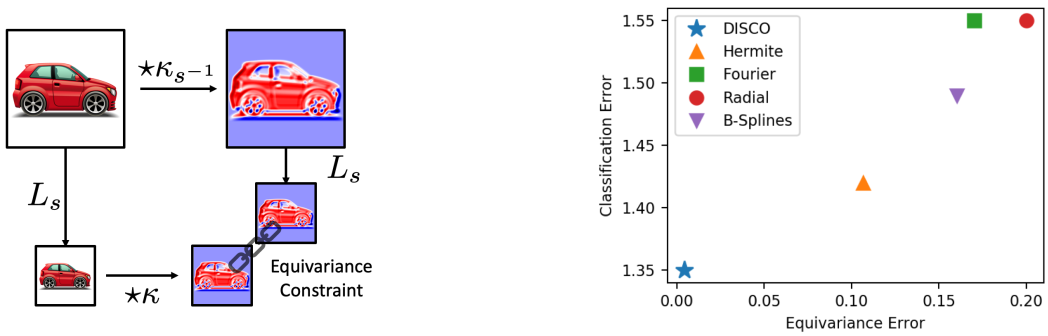

Figure 1. Left: the necessary constraint for scale-equivariance. When it is not satisfied an *equivariance error* appears. Right: Equivariance error vs. Classification error for scale-equivariant models on MNIST-scale. DISCO achieves the lowest equivariance error and this leads to the best classification accuracy. Alongside DISCO, we test SESN models with Hermite [28], Fourier [39], Radial [13] and B-Spline [3] bases.

2. Related Work

Group Equivariant Networks In recent years, various works on group-equivariant convolution neural networks have appeared. In majority, they consider the roto-translation group in 2D [9, 10, 15, 35, 30, 32], the roto-translation group in 3D [33, 17, 29, 6, 31], or the rotation group in 3D [6, 12, 8]. In [7, 18, 19] the authors demonstrate how to build convolution networks equivariant to arbitrary compact groups. All these papers cover group-equivariant networks for compact groups. In this paper, we focus the scale-translation group which is an example of a non-compact group.

Discrete Operators Minimization of the discrepancies between the theoretical properties of continuous models and their discrete realizations has been studied for a variety of computer vision tasks. Lindeberg [21, 22] proposed a method for building a scale-space for discrete signals. The approach relied on the connection between the discretized version of the diffusion equation and the structure of images. While this method considered the scale symmetry of images and significantly improved computer vision models in the pre-deep-learning era, it is not directly applicable to our case of scale-equivariant convolutional networks.

In [11], Diaconu and Worrall demonstrate how to construct rotation-equivariant CNNs on the pixel grid for arbitrary rotations. The authors propose to learn the kernels which minimize the equivariance error of rotation-equivariant convolutional layers. The method relies on the properties of the rotation group and cannot be generalized to the scale-translation group. In this paper, we show how to minimize the equivariance error for scale-convolution without the use of extensive learning.

Scale-Equivariant CNNs An early work of [16] introduced SI-ConvNet, a model where the input image is rescaled into a multi-scale pyramid. Alternatively, Xu *et al.* [37] proposed SiCNN, where a multi-scale representation is built from rescaling the network filters. While these networks significantly improve image classification, they are several orders slower than standard CNNs.

In [28, 3, 39] the authors propose to parameterize the filters by a trainable linear combination of a pre-calculated, fixed multi-scale basis. Such a basis is defined in the continuous scale domain and projected on a pixel grid for the set of scale factors. The models do not involve interpolation during training nor inference. As a consequence, they operate within reasonable time. The continuous nature of the bases allows for the use of arbitrary scale factors, but it suffers from a reduced accuracy as the projection on the discrete grid causes an equivariance error.

Worral and Welling [34] propose to model filter rescaling by dilation. This solves the equivariance error of the previous method at the price of permitting only integer scale factors. That makes the method less suited for object tracking, depth analysis and fine-grained image classification, where subtle changes in the image scale are important in the performance. Our approach combines the best of the both worlds as it guarantees a low equivariance error for arbitrary scale factors.

3. Method

Equivariance A mapping g is equivariant under a transformation L if and only if there exists L' such that $g \circ L = L' \circ g$. If the mapping L' is identity, then g is invariant under transformation L.

Scale Transformations Given a function $f : \mathbb{R} \to \mathbb{R}$ its scale transformation L_s is defined by

$$L_s[f](t) = f(s^{-1}t), \quad \forall s > 0 \tag{1}$$

We refer to cases with $s > 1$ as up-scalings and to cases with $s < 1$ as down-scalings, where $L_{1/2}[f]$ stands for a function down-scaled by a factor of 2.

Model	Basis	MNIST	MNIST+	Equi. error	# Params.
CNN	-	2.02 ± 0.07	1.60 ± 0.09	-	495 K
SiCNN	-	2.02 ± 0.14	1.59 ± 0.03	-	497 K
SI-ConvNet	-	1.82 ± 0.11	1.59 ± 0.10	-	495 K
SEVF	-	2.12 ± 0.13	1.81 ± 0.09	-	475 K
DSS	Dilation	1.97 ± 0.08	1.57 ± 0.09	0.0	494 K
SS-CNN	Radial	1.84 ± 0.10	1.76 ± 0.07	-	494 K
SESN	Hermite	1.68 ± 0.06	1.42 ± 0.07	0.107	495 K
SESN	B-Spline	1.74 ± 0.08	1.49 ± 0.05	0.163	495 K
SESN	Fourier	1.88 ± 0.07	1.55 ± 0.07	0.170	495 K
SESN	Radial	1.74 ± 0.07	1.55 ± 0.10	0.200	495 K
DISCO	Discrete	$\mathbf{1.52 \pm 0.06}$	$\mathbf{1.35 \pm 0.05}$	0.004	495 K

Table 1. The classification error of various methods on the MNIST-scale dataset, lower is better. We test both the regime with and without data augmentation, where scaling data augmentation is denoted by "+". All results are reported as mean $\pm$ std over 6 different, fixed realizations of the dataset. The best results are **bold**.

The scale-translation group We are interested in equivariance under the scale-translation group H and its subgroups. It consists of the translations t and scale transformations s which preserve the position of the center. $H = \{(s,t)\} = S \rtimes T$ is a semi-direct product of a multiplicative group $S = (\mathbb{R}^+, +)$ and an additive group $T = (\mathbb{R}, +)$. For the multiplication of its elements we have $(s_2, t_2) \cdot (s_1, t_1) = (s_1 s_2, s_2 t_1 + t_2)$. Scale transformation of a function defined on group H consists of a scale transformation of its spatial part as it is defined in the Equation 1 and a corresponding multiplicative transformation of its scale part. In other words

$$L_{\hat{s}}[f](s,t) = f(s\hat{s}^{-1}, \hat{s}^{-1}t) \tag{2}$$

3.1. Scale-Convolution

A scale-convolution of f and a kernel κ both defined on scale s and translation t is given by: [28]:

$$[f \star_H \kappa](s,t) = \sum_{s'} [f(s', \cdot) \star \kappa_s(s^{-1}s', \cdot)](\cdot, t) \tag{3}$$

where κ_s stands for an s-times up-scaled kernel κ, $\star$ and $\star_H$ are convolution and scale-convolution. The exact way the up-scaling is performed depends on how the down-scaling of the input signal works.

Scale-convolution is equivariant to transformations $L_{\hat{s}}$ from the group H, therefore the following holds true by definition:

$$[L_{\hat{s}}[f] \star_H \kappa] = L_{\hat{s}}[f \star_H \kappa] \tag{4}$$

Expanding the left and the right hand side of this relation by using Equation 3, choosing $s = 1$ and replacing $s' \to s'\hat{s}$ we find:

$$L_s[f] \star \kappa = L_s[f \star \kappa_{s^{-1}}], \quad \forall f, s \tag{5}$$

The mapping defined by Equation 3 is scale-equivariant only if a kernel and its up-scaled versions satisfy Equation 5. In [28, 3, 39] the opposite, sufficient condition was proved.

3.2. Exact Solution

In the continuous domain, convolution is defined as an integral over the spatial coordinates. [28, 3, 39] derives a solution for Equation 5:

$$\kappa_s(t) = s^{-1}\kappa(s^{-1}t) \tag{6}$$

However, when such kernels are calculated and projected on the pixel grid, a discrepancy between the left-hand side and the right-hand side of Equation 5 will appear. We refer to such inequality as the *equivariance error*.

We aim at directly solving Equation 5 in the discrete domain. In general, for discrete signals down-scaling is a non-invertible operation. Thus L_s is well-defined only for $s < 1$. We start by solving Equation 5 for 1-dimensional discrete signals. The 2-dimensional solution can always be constructed as a linear combination of separable functions. Thus, the relation between these cases is bijective.

Let us consider a discrete signal f represented as a vector $\boldsymbol{f}$ of length N_{in}. It is down-scaled to length $N_{\text{out}} < N_{\text{in}}$ by L_s, which is represented as a rectangular interpolation matrix $\boldsymbol{L}$ of size $N_{\text{out}} \times N_{\text{in}}$. A convolution with a kernel κ is represented as a multiplication with a matrix $\boldsymbol{K}$ of size $N_{\text{out}} \times N_{\text{out}}$, and with a kernel $\kappa_{s^{-1}}$ written as a matrix $\boldsymbol{K}_{s^{-1}}$ of size $N_{\text{in}} \times N_{\text{in}}$. Then Equation 5 can be rewritten in matrix form as follows:

$$\boldsymbol{K}\boldsymbol{L}\boldsymbol{f} = \boldsymbol{L}\boldsymbol{K}_{s^{-1}}\boldsymbol{f}, \ \forall \boldsymbol{f} \quad \iff \quad \boldsymbol{K}\boldsymbol{L} = \boldsymbol{L}\boldsymbol{K}_{s^{-1}} \tag{7}$$

Without loss of generality we assume circular boundary conditions. Then the matrix representations $\boldsymbol{K}$ and $\boldsymbol{K}_{s^{-1}}$

are both circulant and their eigenvectors are the column-vectors of the Discrete Fourier Transform $\boldsymbol{F}$ [2]. The solution with respect to $\boldsymbol{\kappa}_{s^{-1}}$ is the dilation of $\boldsymbol{\kappa}$ by factor s. Such a solution also known as the *à trous algorithm* [14]:

$$(\kappa_{s^{-1}})_{is} = \sum_i \boldsymbol{F}^*_{ij}(\boldsymbol{KLF})_{1j}/(\boldsymbol{LF})_{1j} = \kappa_i \tag{8}$$

3.3. Approximate solution

Let us consider a scale-convolutional layer with a set of scales $\{1, \sqrt{2}, 2, 2\sqrt{2}, 4, 4\sqrt{2}, \dots\}$. The set of corresponding kernels is $\{\kappa_1, \kappa_{\sqrt{2}}, \kappa_2, \kappa_{2\sqrt{2}}, \dots\}$. As the smallest kernel is known, all kernels defined on integer scales can be calculated as its dilated versions. And, when kernel $\kappa_{\sqrt{2}}$ is defined, all intermediate kernels $\kappa_{2\sqrt{2}}, \kappa_{4\sqrt{2}}, \dots$ can be calculated by using dilation as well. Thus, the only kernel yet unknown is kernel $\kappa_{\sqrt{2}}$.

The kernel $\kappa_{\sqrt{2}}$ can be calculated as a minimizer of the equivariance error based on the Equation 5 as follows:

$$\begin{aligned}\kappa_{\sqrt{2}} = \arg\min \mathbb{E}_f \|L[f] \star \kappa_1 - L[f \star \kappa_{\sqrt{2}}]\|^2_F \\ + \|L[f] \star \kappa_{\sqrt{2}} - L[f \star \kappa_2]\|^2_F\end{aligned} \tag{9}$$

where $L = L_{1/\sqrt{2}}$ is a down-scaling by a factor $\sqrt{2}$.

To construct scale-equivariant convolution we parametrize the kernels as a linear combination of a fixed multi-scale basis calculated accroding to Equation 8, Equation 9. The basis is then fixed and only corresponding coefficients are trained. The coefficients are shared for all scales. We refer to scale-convolutions with the proposed bases as Discrete Scale Convolutions or shortly DISCO. As DISCO kernels are sparse, they allow for lower computational complexity.

4. Experiments

4.1. Equivariance Error

To quantitatively evaluate the equivariance error of DISCO versus other methods for scale-convolution [28, 39, 3], we follow the approach proposed in [28]. In particular, we randomly sample images from the MNIST-Scale dataset [28] and pass in through the scale-convolution layer. Then, the equivariance error is calculated as follows:

$$\Delta = \sum_s \|L_s\Phi(f) - \Phi(L_s f)\|^2_2 / \|L_s\Phi(f)\|^2_2 \tag{10}$$

where Φ is scale-convolution with weights initialized randomly.

The equivariance error for each model is reported in Table 1 and in Figure 1. As can be seen, there exists a correlation between an equivariance error and classification accuracy. DISCO model attains the lowest equivariance error.

Model	Basis	STL-10	Time, s
WRN	-	11.48	10
SiCNN	-	11.62	110
SI-ConvNet	-	12.48	55
DSS	Dilation	11.28	40
SS-CNN	Radial	25.47	15
SESN	Hermite	8.51	165
DISCO	Discrete	**8.07**	50

Table 2. The classification error on STL-10. The best results are in **bold**. The average compute time per epoch is reported in seconds.

4.2. Image Classification

Alongside DISCO, we test SI-ConvNet [16], SS-CNN [13], SiCNN [37], SEVF [23], DSS [34] and SESN [28]. We additionally reimplement SESN models with other bases such as B-Splines [3], Fourier-Bessel Functions [39] and Log-Radial Harmonics [13, 24].

MNIST-scale Following [28] we conduct experiments on the MNIST-scale dataset. As a baseline model we use the SESN model, which holds the state-of-the-art result on this dataset. Both SESN and DISCO use the same set of scales in scale convolutions: $\{1, 2^{1/3}, 2^{2/3}, 2\}$ and are trained in exactly the same way. As can be seen from Table 1, our DISCO model outperforms other scale equivariant networks in accuracy and equivariance error and sets a new state-of-the-art result.

STL-10 To demonstrate how accurate scale equivariance helps when the training data is limited, we conduct experiments on the STL-10 [5] dataset. As a baseline we use WideResNet [38] with 16 layers and a widening factor of 8. Scale-equivariant models are constructed according to [28]. All models have the same number of parameters, the same set of scales $\{1, \sqrt{2}, 2\}$ and are trained for the same number of steps.

As can be seen from Table 2, the proposed DISCO model outperforms the other scale-equivariant networks and sets a new state-of-the-art result in the supervised learning setting. Moreover, DISCO is more than 3 times faster than the second-best SESN-model.

5. Discussion

In this work, we demonstrate that the equivariance error affects the performance of equivariant networks. We introduce DISCO, an approach to rescale a basis in scale-convolution, so the equivariance error is minimized. We experimentally demonstrate that DISCO scale-equivariant networks outperform conventional and other scale-equivariant models, setting the new state-of-the-art MNIST-Scale and improving results on STL-10 datasets.

References

[1] Helen L Anderson, Ruzena Bajcsy, and Max Mintz. Adaptive image segmentation. 1988. 1

[2] Bassam Bamieh. Discovering transforms: A tutorial on circulant matrices, circular convolution, and the discrete fourier transform. *arXiv preprint arXiv:1805.05533*, 2018. 4

[3] Erik J Bekkers. B-spline cnns on lie groups. *arXiv preprint arXiv:1909.12057*, 2019. 1, 2, 3, 4

[4] Jason Chang and John W Fisher. Analysis of orientation and scale in smoothly varying textures. In *2009 IEEE 12th International Conference on Computer Vision*, pages 881–888. IEEE, 2009. 1

[5] Adam Coates, Andrew Ng, and Honglak Lee. An analysis of single-layer networks in unsupervised feature learning. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 215–223, 2011. 4

[6] Taco Cohen, Mario Geiger, Jonas Köhler, and Max Welling. Convolutional networks for spherical signals. *arXiv preprint arXiv:1709.04893*, 2017. 2

[7] Taco Cohen, Mario Geiger, and Maurice Weiler. A general theory of equivariant cnns on homogeneous spaces. *arXiv preprint arXiv:1811.02017*, 2018. 2

[8] Taco Cohen, Maurice Weiler, Berkay Kicanaoglu, and Max Welling. Gauge equivariant convolutional networks and the icosahedral cnn. In *International Conference on Machine Learning*, pages 1321–1330. PMLR, 2019. 2

[9] Taco Cohen and Max Welling. Group equivariant convolutional networks. In *International conference on machine learning*, pages 2990–2999, 2016. 1, 2

[10] Taco S Cohen and Max Welling. Steerable cnns. *arXiv preprint arXiv:1612.08498*, 2016. 2

[11] Nichita Diaconu and Daniel Worrall. Learning to convolve: A generalized weight-tying approach. In *International Conference on Machine Learning*, pages 1586–1595. PMLR, 2019. 2

[12] Carlos Esteves, Christine Allen-Blanchette, Ameesh Makadia, and Kostas Daniilidis. Learning so (3) equivariant representations with spherical cnns. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 52–68, 2018. 2

[13] Rohan Ghosh and Anupam K Gupta. Scale steerable filters for locally scale-invariant convolutional neural networks. *arXiv preprint arXiv:1906.03861*, 2019. 2, 4

[14] Matthias Holschneider, Richard Kronland-Martinet, Jean Morlet, and Ph Tchamitchian. A real-time algorithm for signal analysis with the help of the wavelet transform. In *Wavelets*, pages 286–297. Springer, 1990. 4

[15] Emiel Hoogeboom, Jorn WT Peters, Taco S Cohen, and Max Welling. Hexaconv. *arXiv preprint arXiv:1803.02108*, 2018. 2

[16] Angjoo Kanazawa, Abhishek Sharma, and David Jacobs. Locally scale-invariant convolutional neural networks. *arXiv preprint arXiv:1412.5104*, 2014. 1, 2, 4

[17] Risi Kondor. N-body networks: a covariant hierarchical neural network architecture for learning atomic potentials. *arXiv preprint arXiv:1803.01588*, 2018. 2

[18] Risi Kondor and Shubhendu Trivedi. On the generalization of equivariance and convolution in neural networks to the action of compact groups. In *International Conference on Machine Learning*, pages 2747–2755. PMLR, 2018. 2

[19] Leon Lang and Maurice Weiler. A wigner-eckart theorem for group equivariant convolution kernels. *arXiv preprint arXiv:2010.10952*, 2020. 2

[20] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2117–2125, 2017. 1

[21] Tony Lindeberg. Scale-space for discrete signals. *IEEE transactions on pattern analysis and machine intelligence*, 12(3):234–254, 1990. 2

[22] Tony Lindeberg. *Scale-space theory in computer vision*, volume 256. Springer Science & Business Media, 2013. 2

[23] Diego Marcos, Benjamin Kellenberger, Sylvain Lobry, and Devis Tuia. Scale equivariance in cnns with vector fields. *arXiv preprint arXiv:1807.11783*, 2018. 4

[24] Hanieh Naderi, Leili Goli, and Shohreh Kasaei. Scale equivariant cnns with scale steerable filters. In *2020 International Conference on Machine Vision and Image Processing (MVIP)*, pages 1–5. IEEE, 2020. 4

[25] Henry Schneiderman and Takeo Kanade. Object detection using the statistics of parts. *International Journal of Computer Vision*, 56(3):151–177, 2004. 1

[26] Arnold WM Smeulders, Dung M Chu, Rita Cucchiara, Simone Calderara, Afshin Dehghan, and Mubarak Shah. Visual tracking: An experimental survey. *IEEE transactions on pattern analysis and machine intelligence*, 36(7):1442–1468, 2013. 1

[27] Ivan Sosnovik, Artem Moskalev, and Arnold W.M. Smeulders. Scale equivariance improves siamese tracking. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 2765–2774, January 2021. 1

[28] Ivan Sosnovik, Michał Szmaja, and Arnold Smeulders. Scale-equivariant steerable networks. *arXiv preprint arXiv:1910.11093*, 2019. 1, 2, 3, 4

[29] Nathaniel Thomas, Tess Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation-and translation-equivariant neural networks for 3d point clouds. *arXiv preprint arXiv:1802.08219*, 2018. 2

[30] Maurice Weiler and Gabriele Cesa. General $e(2)$-equivariant steerable cnns. *arXiv preprint arXiv:1911.08251*, 2019. 2

[31] Maurice Weiler, Mario Geiger, Max Welling, Wouter Boomsma, and Taco Cohen. 3d steerable cnns: Learning rotationally equivariant features in volumetric data. In *Advances in Neural Information Processing Systems*, pages 10381–10392, 2018. 2

[32] Maurice Weiler, Fred A Hamprecht, and Martin Storath. Learning steerable filters for rotation equivariant cnns. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 849–858, 2018. 2

[33] Daniel Worrall and Gabriel Brostow. Cubenet: Equivariance to 3d rotation and translation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 567–584, 2018. 2

[34] Daniel Worrall and Max Welling. Deep scale-spaces: Equivariance over scale. In *Advances in Neural Information Processing Systems*, pages 7366–7378, 2019. 1, 2, 4

[35] Daniel E Worrall, Stephan J Garbin, Daniyar Turmukhambetov, and Gabriel J Brostow. Harmonic networks: Deep translation and rotation equivariance. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5028–5037, 2017. 2

[36] Yiran Wu, Sihao Ying, and Lianmin Zheng. Size-to-depth: a new perspective for single image depth estimation. *arXiv preprint arXiv:1801.04461*, 2018. 1

[37] Yichong Xu, Tianjun Xiao, Jiaxing Zhang, Kuiyuan Yang, and Zheng Zhang. Scale-invariant convolutional neural networks. *arXiv preprint arXiv:1411.6369*, 2014. 2, 4

[38] Sergey Zagoruyko and Nikos Komodakis. Wide residual networks. In *BMVC*, 2016. 4

[39] Wei Zhu, Qiang Qiu, Robert Calderbank, Guillermo Sapiro, and Xiuyuan Cheng. Scale-equivariant neural networks with decomposed convolutional filters. *arXiv preprint arXiv:1909.11193*, 2019. 2, 3, 4