UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

# Geometric Back-Propagation in Morphological Neural Networks

Groenendijk, R.; Dorst, L.; Gevers, T.

DOI
10.1109/TPAMI.2023.3290615

Publication date
2023

Document Version
Final published version

Published in
IEEE Transactions on Pattern Analysis and Machine Intelligence

License
CC BY

Link to publication

Citation for published version (APA):
Groenendijk, R., Dorst, L., & Gevers, T. (2023). Geometric Back-Propagation in Morphological Neural Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(11), 14045-14051. https://doi.org/10.1109/TPAMI.2023.3290615

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

*UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)*

Download date:14 Nov 2026

# Geometric Back-Propagation in Morphological Neural Networks

Rick Groenendijk, Leo Dorst, and Theo Gevers

***Abstract*—This paper provides a definition of back-propagation through geometric correspondences for morphological neural networks. In addition, dilation layers are shown to learn probe geometry by erosion of layer inputs and outputs. A proof-of-principle is provided, in which predictions and convergence of morphological networks significantly outperform convolutional networks.**

***Index Terms*—Mathematical morphology, morphological neural networks, back-propagation, probe geometry, depth infilling.**

## I. Introduction

There exists much image-like data that is produced by contact probing: examples are depth by LiDAR or from time-of-flight sensors, radar images, and scanning microscopy. Since Serra's investigations [1], [2], the underlying algebraic structure of such data has been known as *mathematical morphology*. A coherent set of operations forms a consistent alternative framework to the convolution way of linear diffusion probing [3], [4], [5]. The latter forms the basis of the convolutional neural network (CNN); and people have wondered about developing analogous morphological neural networks (MNNs) to process the morphological type of data [6], [7], [8]. Convolution does not inherently respect the separation between pixels in 3D space, treating them as equidistant neighbors at all times, and cannot process occlusion naturally. On the other hand, morphological operations such as dilation, allow data to be probed by structuring elements in space, respecting separation and occlusion.

MNNs have recently seen a variety of successes in complex vision tasks: [9], [10] show that MNNs have vastly higher parameter efficiency in tasks such as digit recognition; [11] successfully removes artefacts in images caused by rain droplets and their method is extended in [12] introducing opening-closing networks; [13] shows that morphological operations and convolutions can supplement each other and achieve state-of-the-art performance in object boundary recognition; [14] uses deep MNNs to solve classification and multi-class segmentation tasks. The authors note that training becomes increasingly challenging as networks are deepened with more complex topology; [15] extends the work of [16] on equivariant scale networks by the use of morphological scale spaces, though only the first module of their network is actually morphological.

Even outside the scope of neural networks, morphology is used to encode rich computational features that are useful in a variety of contact-related tasks: [17] encodes surfaces of fractured archaeological objects into a set of morphological features with the goal of automatically fitting fragments; [18] develops a morphological variant of Deep Mode Composition that performs 1D image segmentation by the use of a morphological-convolution hybrid.

Neural networks are trained using back-propagation to minimize an objective function. Morphological operations are not fully differentiable however; therefore MNNs cannot readily be trained using linear back-propagation. As a consequence, three different approaches to back-propagating errors over sequences of morphological filters are established in the literature: *(1)* approximate (linearly) the *min, max*-operations to make them differentiable [10], [19], [20]; *(2)* use a sub-gradient definition in which the error is only propagated over *min, max*-elements – similar to derivatives over pooling operations for spatial sub-sampling in deep learning frameworks [21]; and *(3)* use depth- and point-wise convolutions in conjunction with pooling operations [14], [22]. Optimization-based updating schemes, e.g., the Convex-Concave procedure in [8], are beyond the scope of this article. Most often *(2)* is used because it is facilitated by modern deep learning frameworks – see e.g., [9], [11], [13], [23]. For any of these methods, stable learning and convergence is not guaranteed. It will be shown that all these methods can be improved by employing a principled morphological back-propagation. The theoretical aim of this paper is therefore to derive a geometric interpretation of morphological back-propagation and formulate an alternate updating rule for learning probe geometry by Morphological Neural Networks.

The contributions of this paper are:

- A geometric definition of the back-propagation of morphological operations that does not rely on linear approximation of morphological operations as previous works did, but rather on matching slopes of (locally convex) functions.
- A morphological definition for probe geometry learning by error bounding, especially suited to data acquired by an essentially morphological process.
- Confirmation of the theory in practical use on probe geometry estimation in Scanning Probe Microscopy (SPM) and depth infilling on NYUv2.

## II. Method

The goal of this paper is the direct application of morphological operations to neural networks and to back-propagate errors during training. As a brief recap, neural networks use the back-propagation algorithm [24] to update parameters and approximate a function for which data samples are available. Consider a network built up of $L$ layers as a composite function $f(\mathbf{x}) = f_L \circ f_{L-1} \circ \ldots f_1(\mathbf{x})$. Consider also a (sample, target) dataset $\mathbf{D} = \{(\mathbf{x}_n, \mathbf{t}_n)\}_{n=0}^{N}$ and let $f(\mathbf{x}_n) = \mathbf{y}_n$ be the output of the network. The network is trained through minimizing a number of training objectives $E(\mathbf{t}_n, \mathbf{y}_n)$, where $\mathbf{t}_n$ is the target output. The advantage of the back-propagation algorithm is that updating the parameters of a single layer can be agnostic of the network architecture, as long as the local derivative of the error $\frac{\partial E}{\partial f_+(\mathbf{y})}$ is known at the required point. For the remainder of this paper, a per-layer notation is therefore used, with input $f_-$, output $f_+$, probe $p$, and corresponding derivatives, as in Fig. 1. For MNNs, the terms $\frac{\partial f_+}{\partial f_-}$, $\frac{\partial f_+}{\partial p}$ have to be redefined because they differ from those in CNNs.

Manuscript received 17 July 2022; revised 26 May 2023; accepted 23 June 2023. Date of publication 29 June 2023; date of current version 3 October 2023. This work was supported by Dutch Research Council (NWO) through project FlexCRAFT under Grant P17-01. Recommended for acceptance by J.M. Solomon. *(Corresponding author: Rick Groenendijk.)*

The authors are with the Computer Vision Group, University of Amsterdam, 1012 WP Amsterdam, The Netherlands (e-mail: r.w.groenendijk@uva.nl; l.dorst@uva.nl; th.gevers@uva.nl).

Digital Object Identifier 10.1109/TPAMI.2023.3290615

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

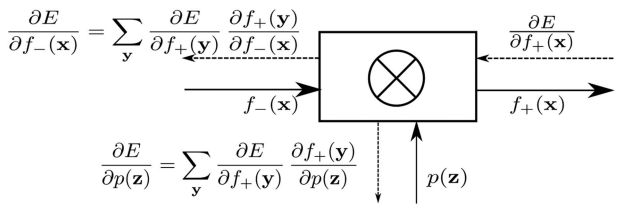


Fig. 1. Schematic overview of a single layer in a neural network. It shows forward propagation of $f_-$ and $p$ by some operation $\otimes$, and backward propagation of the error $\frac{\partial E}{\partial f_+(\mathbf{x})}$. From the chain rule, only the terms $\frac{\partial E}{\partial f_+(\mathbf{y})}$ are required to obtain the derivatives of the error with respect to the input $f_-$ and the parameterized probe $p$ (or kernel) of $\otimes$.

Dilation for functions is defined on the semi-ring $(\mathbb{R}_{-\infty}, \vee, +)$ where $\vee$ denotes the supremum operation and $+$ is addition. This algebraic system extends the set of reals $\mathbb{R}$ with minus infinity: $\mathbb{R}_{-\infty} \equiv \mathbb{R} \cup -\infty$. In MNNs, the multiplication-addition scheme of convolution is replaced by an addition-supremum scheme of dilation [6].

A layer input signal $f_- : \mathbb{R}^D \to \mathbb{R}_{-\infty}$ indexed by indicator variable $\mathbf{x}$, and a structuring element (or probe) $p : \mathbb{R}^D \to \mathbb{R}_{-\infty}$ indexed by indicator variable $\mathbf{z}$ are combined to produce the morphological dilation as the layer output signal:

$$f_+(\mathbf{x}) = \bigvee_{\mathbf{z}} f_-(\mathbf{x} - \mathbf{z}) + p(\mathbf{z}). \tag{1}$$

Morphological back-propagation is derived in Section II-C through slope correspondences explained in Section II-A. As an intermediate step, the morphological derivative is given in Section II-B. Finally, in Section II-E, elements that do not contribute to any output in the forward pass, are bounded by means of morphological erosion in the backward pass of back-propagation. For brevity, a dilation layer is used in all derivations, but by morphological duality [1], [2] all arguments can be made for an erosion layer as well.

### A. Slope Correspondences

Geometrically speaking, it is intuitive to regard morphological dilation as probing a signal with a mirrored and flipped structuring element $p^T(\mathbf{z}) \equiv -p(-\mathbf{z})$ from above, lowering it until there is at least a single point of contact. As the probe moves, the reference point of the probe $p^T$ traces out the output signal $f_+$. There may be several points of contact or even entire ranges where $f_-$ and $p^T$ can be in touch. On the other hand, it is never allowed that $p^T$ and $f_-$ intersect. A graphical example is shown in Fig. 2.

Not all locations $\mathbf{x}_-$ on the input signal $f_-$ lead to an output point $(\mathbf{x}_+, f_+(\mathbf{x}_+))$ since not all $f_-(\mathbf{x}_-)$ can be touched by $p^T$. However, all $\mathbf{x}_+$ can be traced back to at least one location $\mathbf{x}_-$. For *back*-propagation of the error –as required for network learning– the goal is to map an error at $\mathbf{x}_+$ back to *any* $\mathbf{x}_-$ that caused it.

*Theorem 1 (Originally From [3]).* There is a provenance relationship between the slope of any point on the output signal $f_+(\mathbf{x}_+)$, and the points that caused it through contact of the input signal $f_-(\mathbf{x}_-)$ and the probe $p(\mathbf{z}_-)$. The contact location points are related through:

$$\mathbf{x}_+ = \mathbf{x}_- + \mathbf{z}_-, \tag{2}$$

and the slopes (gradients) obey

$$\nabla f_+(\mathbf{x}_+) = \nabla f_-(\mathbf{x}_-) = \nabla p(\mathbf{z}_-). \tag{3}$$

*Proof:* For a point $(\mathbf{x}_+, f_+(\mathbf{x}_+))$, there is always at least one input location $\mathbf{x}_-$ at the input signal $f_-$ for which the supremum $\bigvee_{\mathbf{z}} f_-(\mathbf{x} - \mathbf{z}) + p(\mathbf{z})$ is attained. The contact location $\mathbf{x}_-$ implies the existence of a location $\mathbf{z}_-$ on the probe $p$ that satisfies $\mathbf{z}_- = \mathbf{x}_+ - \mathbf{x}_-$ where $\mathbf{z}_- \in \mathrm{dom}(p)$. Therefore, (1) can be rewritten in terms of the locations on $f_+$, $f_-$, and $p$ where the supremum occurs:

$$f_+(\mathbf{x}_+) = f_-(\mathbf{x}_+ - \mathbf{z}_-) + p(\mathbf{z}_-). \tag{4}$$

More specifically, the supremum is attained when the first derivative of $f_+$ with respect to $\mathbf{z}_-$ at $\mathbf{x}_+$ is zero (there is also a second order condition to make it a supremum rather than an infimum). As a consequence:

$$\nabla_{\mathbf{z}_-}\left(f_-(\mathbf{x}_+ - \mathbf{z}_-) + p(\mathbf{z}_-)\right) = 0. \tag{5}$$

The $f_-$-slope at $\mathbf{x}_-$ now relates to the $p$-slope at $\mathbf{z}_-$:

$$\begin{aligned}\nabla_{\mathbf{x}_-} f_-(\mathbf{x}_-) &= [\nabla_{\mathbf{x}_-}\mathbf{z}_-]\ \nabla_{\mathbf{z}_-} f_-(\mathbf{x}_+ - \mathbf{z}_-) \\ &= [\nabla_{\mathbf{x}_-}(\mathbf{x}_+ - \mathbf{x}_-)]\,(-\nabla_{\mathbf{z}_-} p(\mathbf{z}_-)), \\ &= \nabla_{\mathbf{z}_-} p(\mathbf{z}_-).\end{aligned} \tag{6}$$

At contact, those slopes are also related to the slope of the output signal $f_+$, by differentiating (4) to $\mathbf{x}_+$:

$$\nabla_{\mathbf{x}_+} f_+(\mathbf{x}_+) = \nabla_{\mathbf{x}_+} f_-(\mathbf{x}_+ - \mathbf{z}_-). \tag{7}$$

Combining (6) and (7) with the provenance (4) completes the proof.

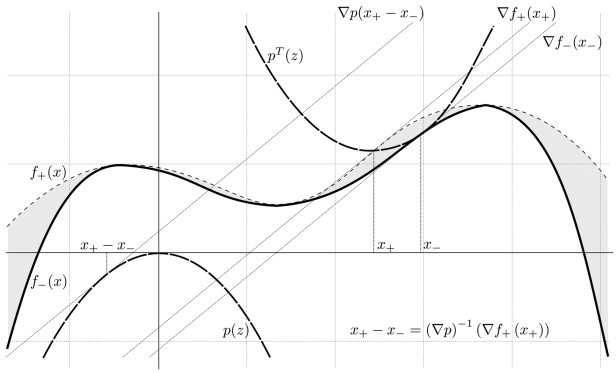


Fig. 2. Depiction of the forward propagation of a 1D signal $f_-$ using morphological dilation as a single-layer of a larger network. The dilation of a signal $f$ with (convex) probe $p$ is shown resulting in $f_+$. Clearly here $\nabla f_+(\mathbf{x}_+) = \nabla f_-(\mathbf{x}_-) = \nabla p(\mathbf{z}_-)$ which is proven in the main text.

### B. Morphological Derivatives

During back-propagation all $(\mathbf{x}_+, f_+(\mathbf{x}_+))$ are known, but their corresponding $\mathbf{x}_-$ must be determined. These correspondences, henceforth called the *provenance* of points, can be obtained by matching slopes using probe $p$.

*Theorem 2:* The morphological derivative of a single-layer $f_+$ with respect to the input $f_-$ is

$$\frac{\partial f_+(\mathbf{x}_+)}{\partial f_-}(\mathbf{x}_-) = \begin{cases} 1 & \forall\left[\nabla p(\mathbf{x}_+ - \mathbf{x}_-) = \nabla f_+(\mathbf{x}_+)\right] \\ 0 & \text{otherwise,} \end{cases} \tag{8}$$

where $\forall[\ldots]$ denotes the set of points $(\mathbf{x}_+, \mathbf{x}_-)$ for which the equality holds, and $\frac{\partial f_-(\mathbf{x}_+)}{\partial f_-}(\mathbf{x}_-)$ denotes the derivative of $f_+(\mathbf{x}_+)$ with respect to $f_-$ evaluated at $\mathbf{x}_-$, so $\frac{\partial f_+(\mathbf{x}_+)}{\partial f_-(\mathbf{x}_-)}$ as a function of $\mathbf{x}_-$.

*Proof:* The derivative of $f_+(\mathbf{x}_+)$ in (1) with respect to $f_-$ is 1 if and only if $f_-(\mathbf{x}_-)$ caused $f_+(\mathbf{x}_+)$. In any other case, the derivative is zero. Moreover, as a consequence of Theorem 1 each layer output location $\mathbf{x}_+$ relates to a location $\mathbf{x}_-$ on the input layer $f_-$ that caused the corresponding $f_+(\mathbf{x}_+)$ given the current $p$ by means of $\nabla p(\mathbf{x}_+ - \mathbf{x}_-) = \nabla f_+(\mathbf{x}_+)$.

There are four subtleties captured in Theorem 2 that are not immediately apparent. The first concerns undefined provenances; the latter three concern the *for all* (i.e., $\forall[\dots]$) statement.

*Undefined Provenance:* There may be locations $\mathbf{x}_-$ that have caused not a single $\mathbf{x}_+$ in the forward pass, resulting in a derivative in the backward pass that is zero for that $\mathbf{x}_-$. At those locations $\mathbf{x}_-$ the provenance (i.e., the correspondence between $\mathbf{x}_-$ and $\mathbf{x}_+$) is undefined. Therefore, (8) is called a *sub-gradient* [10], since it is not a local rate of change with any $\mathbf{x}_-$, but rather a zero-valued derivative resulting from an undefined provenance between $\mathbf{x}_+$ to $\mathbf{x}_-$.

*Multiple $\mathbf{x}_+$, Single $\mathbf{x}_-$:* Multiple $\mathbf{x}_+$ may have been caused by a single $\mathbf{x}_-$. Equation (8) allows $\frac{\partial f_+(\mathbf{x}_+)}{\partial f_-}$ to be back-propagated to a single $\mathbf{x}_-$ even when it had caused multiple $\mathbf{x}_+$ in the forward pass; the derivative $\frac{\partial f_+(\mathbf{x}_+)}{\partial f_-}(\mathbf{x}_-)$ is 1 when there exists at least one pair for which the equality $\nabla p(\mathbf{x}_+ - \mathbf{x}_-) = \nabla f_+(\mathbf{x}_+)$ holds.

*Single $\mathbf{x}_+$, Multiple $\mathbf{x}_-$:* There may be an $f_+(\mathbf{x}_+)$ caused by multiple $\mathbf{x}_-$, then $f_+$ at $\mathbf{x}_+$ is not differentiable [25]. At these singular points a one-sided derivative (e.g., left and right-sided derivatives in 1D) needs to be used to obtain valid slopes matching an $\mathbf{x}_+$ to each $\mathbf{x}_-$.

*Invertible $\nabla p$.* For a strictly convex probe, the location $\mathbf{x}_-$ can directly be inferred from matching the slopes of the probe $p$ and the output signal $f_+$ since then $\nabla p$ is invertible. Using (3), the invertibility of $p$ due to convexity, and isolating $\mathbf{x}_-$:

$$\mathbf{x}_- = \mathbf{x}_+ - (\nabla p)^{-1}\left(\nabla f_+\left(\mathbf{x}_+\right)\right), \tag{9}$$

which for a convex $p$ implies

$$\frac{\partial f_+\left(\mathbf{x}_+\right)}{\partial f_-}\left(\mathbf{x}_-\right) = \begin{cases} 1 & \text{if } \mathbf{x}_+ - (\nabla p)^{-1}\left(\nabla f_+\left(\mathbf{x}_+\right)\right) = \mathbf{x}_- \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

### C. Morphological Back-Propagation

With these subtleties noted, the goal is now to define $\frac{\partial E}{\partial f_-}(\mathbf{x}_-)$. The derivative of the error with respect to the input signal $f_-$ is more complicated than the morphological derivatives because the derivative of $E$ with respect to the output layer $f_+$ may be different at each $\mathbf{x}_+$ caused by a single $\mathbf{x}_-$. Since sets of distinct $\frac{\partial E}{\partial f_+}(\mathbf{x}_+)$ cannot be back-propagated to a single $\mathbf{x}_-$, they have to be aggregated. Dilation acts as an absolute effect on its input; a morphological error thus acts absolutely on the terms, not relative to their magnitudes as in convolution. Therefore, only the worst case error should back-propagate. To facilitate this, let $\mathrm{EV} : \mathbb{R}_{-\infty}^{D} \times \mathbb{R} \to \mathbb{R}_{-\infty}$ return the signed most extreme value of a function $f$ over a subset of the domain $\mathbf{x} \subseteq \mathrm{dom}(f)$:

$$\mathrm{EV}\left(f, \mathbf{x}\right) = \begin{cases} \bigvee\limits_{\mathbf{x}} f\left(\mathbf{x}\right) & \text{if } \left|\bigvee\limits_{\mathbf{x}} f(\mathbf{x})\right| \geq \left|\bigwedge\limits_{\mathbf{x}} f\left(\mathbf{x}\right)\right| \\ \bigwedge\limits_{\mathbf{x}} f\left(\mathbf{x}\right) & \text{otherwise}, \end{cases} \tag{11}$$

where $\bigwedge$ denotes the infimum operation. Using the signed most extreme value function EV and combining it with Theorem 2 provides:

$$\frac{\partial E}{\partial f_-}(\mathbf{x}_-) = \begin{cases} \mathrm{EV}\left(\frac{\partial E}{\partial f_+}, \mathbf{x}_+\right) & \forall\left[\nabla p\left(\mathbf{x}_+ - \mathbf{x}_-\right) = \nabla f_+\left(\mathbf{x}_+\right)\right] \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

In summary, the back-propagated error $\frac{\partial E}{\partial f_-}(\mathbf{x}_-)$ is the transfer of the worst case positive or negative $\frac{\partial E}{\partial f_+}(\mathbf{x}_+)$ from the locations $\mathbf{x}_+$ to the location(s) $\mathbf{x}_-$ that caused $\mathbf{x}_+$; these locations are found by matching slopes of the output signal $f_+$ and probe $p$, i.e., through their provenance. Parameters of the probe are updated using gradient descent.

The derivative of the error in the input $f_+$ due to the structuring element $p$ is obtained similarly, since the dilation of (1) is symmetric in $f_-$ and $p$. Observe now that the back-propagated transfer of the error is to the provenance vector $\mathbf{z}_- = \mathbf{x}_+ - \mathbf{x}_-$. Therefore the derivative of the error with respect to $p(\mathbf{z})$ can be written as

$$\frac{\partial E}{\partial p}\left(\mathbf{z}\right) = \begin{cases} \mathrm{EV}\left(\frac{\partial E}{\partial f_+}, \mathbf{x}_+\right) & \forall\left[\nabla p\left(\mathbf{z}\right) = \nabla f_+\left(\mathbf{x}_+\right)\right] \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

In practice, for discrete structures such as images, the (relative) provenance $\mathbf{z}_-$ has to be recorded for each $\mathbf{x}_+$. This bookkeeping can be memory-intensive since an arbitrary number of $\mathbf{x}_-$ may have caused any one $\mathbf{x}_+$ due to the multi-valued nature of dilation, but it prevents issues in approximating slopes from sampled data. Conversely, for data that can reasonably be expected to be locally convex –like SPM– slope matching can theoretically yield a much higher quality gradient with sub-pixel accuracy.

### D. Mean Back-Propagation

It is worth noting that auto-differentiation tools, such as PyTorch [26], use another aggregation of linear back-propagation over multi-valued dilations (and similarly for max-pooling):

$$\frac{\partial E}{\partial f_-}\left(\mathbf{x}_-\right) = \begin{cases} \sum_{\mathbf{x}_+}\left(\frac{1}{|\mathbf{x}_-|}\frac{\partial E}{\partial f_-}\left(\mathbf{x}_+\right)\right) & \text{if } \mathbf{x}_+ - \mathbf{x}_- = \mathbf{z}_- \\ 0 & \text{otherwise,} \end{cases} \tag{14}$$

where $|\mathbf{x}_-|$ is the number of locations $\mathrm{x}_-$ for which the maximum occurred.

The error term is averaged over multiple $\mathbf{x}_-$ that caused a single $\mathbf{x}_+$ to deal with the multi-valued nature of dilation. Then, all $\frac{\partial E}{\partial f_-}\left(\mathbf{x}_+\right)$ are summed for all $\mathbf{x}_+$ that are caused by $\mathbf{x}_-$ to yield a single scalar. In the present context, averaging over the provenance, denoted by $\frac{1}{|\mathbf{x}_-|}$ in (14), is considered to be a non-morphological operation and therefore it is avoided.

It can be surmised, however, that there are practical advantages especially when networks are composed mainly of (linear) convolutions.

### E. Probe Learning by Error Bounding

The main shortcoming of sub-gradient-based morphological back-propagation is that it fails to propagate information about the elements that did *not* cause the error, yielding undefined provenances at those locations. In morphological back-propagation, undefined provenance results in a zero-valued derivative, but there may be better approximations of the error by bounding: since there are points that did not cause the supremum in the forward pass, it means that the error term propagated to those points in (1) can be upper bounded by the points that did cause the supremum.

To see how, consider a dilation with input $f_-$, output $f_+$, and structuring element $p$ as before. The objective of a single layer in the network is to output a $f_+^*$ such that the composite function ultimately minimizes a difference function $Q(\mathbf{x}) = f_L(\mathbf{x}) - t(\mathbf{x})$, where $t(\mathbf{x})$ is the target function at $\mathbf{x}$. Note here that where the morphological back-propagation from Section II-C was agnostic to the form of the error function $E$, the function $Q$ is purely a difference function used to infer $f_+^*$ from $t$.

The input $f_-$ and the ideal output $f_+^*$ for the layer considered are related through an unknown optimal probe:

$$f_+{}^*(\mathbf{x}) \geq \bigvee_{\mathbf{z}} f_-\left(\mathbf{x} - \mathbf{z}\right) + p^*(\mathbf{z}), \tag{15}$$

where $p^*$ is the (optimal) probe to achieve $f_+^*$ from $f_-$. This equation is used to bound $f_+^*$ from below by $\bigvee_{\mathbf{z}} f_-(\mathbf{x} - \mathbf{z}) + p^*(\mathbf{z})$, even when no probe $p^*$ exists to construct $f_+^*$ from $f_-$ or when data is sparse. For a particular $\mathbf{x}' \in \mathbf{x}$, the upper bound output $f_+{}^*$ is given by:

$$f_+{}^*(\mathbf{x}') \geq f_-\left(\mathbf{x}' - \mathbf{z}\right) + p^*(\mathbf{z}), \tag{16}$$

and this in turn implies a bound on $p^*$:

$$p^*(\mathbf{z}) \leq {f_+}^*(\mathbf{y}' + \mathbf{z}) - f_-(\mathbf{y}'), \quad \text{while } \mathbf{y}' = \mathbf{x}' - \mathbf{z}. \tag{17}$$

Letting $\mathbf{x}'$ and hence $\mathbf{y}'$ take all possible values, (17) can be written as an erosion:

$$p^*(\mathbf{z}) \leq \bigwedge_{\mathbf{y}} {f_+}^*(\mathbf{y} + \mathbf{z}) - f_-(\mathbf{y}) = \left(f_+^* \ominus f_-\right)(\mathbf{z}). \tag{18}$$

Therefore, the optimal probe $p^*(\mathbf{z})$ is bounded by the erosion of the desired output ${f_+}^*$ with the input $f_-$.

Similarly to the update rule in gradient descent for convolutional networks, the structuring element $p$ is updated over iterations $i$. Let $p_i$ denote the structuring element at a particular iteration. The update rule for the parameters of $p$ can be given as

$$p_{i+1}(\mathbf{z}) = p_i(\mathbf{z}) - \lambda \Delta p_i(\mathbf{z}) \quad \text{with } \Delta p_i(\mathbf{z}) = p_i([\mathbf{z}) - p_i^*(\mathbf{z}), \tag{19}$$

where $\lambda$ is a gain parameter. This method of learning by bounding will be referred to as *Probe Learning*.

## III. Proof-of-Principle

As a proof-of-principle, this paper shows that when the input data to the MNN is suitable to morphological operations, networks trained by the proposed Morphological Back-propagation and Probe Learning outperform any convolutional network by a large margin. The modality of choice is data resulting from the imaging process of scanning probe microscopy (SPM). The working principle of SPM is positioning a probing element above a surface sample and maintaining constant force in the subclass *atomic force microscopy* (AFM) or constant current in the subclass *scanning tunneling microscopy* (STM). These surface-probe interactions are naturally expressed through mathematical morphology [27], [28], [29], [30], since that is the mathematics of touch probing [4] rather than kernel-based diffusion.

Consider an atomic surface function $S : \mathbb{R}^2 \to \mathbb{R}$ and manufactured probe function $p : \mathbb{R}^2 \to \mathbb{R}$. The resulting image function $I$ obtained from SPM is the morphological dilation of the surface $S$ with the probe $p$:

$$I(\mathbf{x}) = (S \oplus p)(\mathbf{x}) = \bigvee_{\mathbf{z}} S(\mathbf{x} - \mathbf{z}) + p(\mathbf{z}), \tag{20}$$

where $\mathbf{x}, \mathbf{z}$ index spatial locations on the sampling plane for the surface and probe respectively, and $\bigvee$ is the supremum operation. If the geometry of any two of $I, S$, or $p$ is known, the third is related by morphological dilation or erosion. Even when only the scanned image $I$ is known, blind reconstruction techniques [30], [31], [32] may be used to recover an upper bound of the probe $p$ and surface $S$. For a graphical overview, see Fig. 3. In view of (20), data from SPM should have excellent characteristics for testing morphological networks.

In this section, Morphological Back-propagation and Probe Learning are evaluated on synthetic SPM data. The purpose is twofold: to validate the theoretical insights and to show that for appropriate data it is indeed advantageous to incorporate morphological operations in neural networks.

### A. Background

Deep learning is commonly applied in automated analysis of SPM data [33], [34]. Even though it was established decades ago that mathematical morphology can model probe-surface interactions leading to the scanned images (see e.g., [28], [30], [31]), the default operation in automated analysis is still the convolution operator. One argument for using CNNs is that the SPM imaging process is not fully described using morphology: additional noise may be introduced through variance in the tunneling gap in STM or slight cantilever oscillation in AFM [27]. This can partially be modeled by additive Gaussian noise [31], which is inherently difficult for morphology to process.

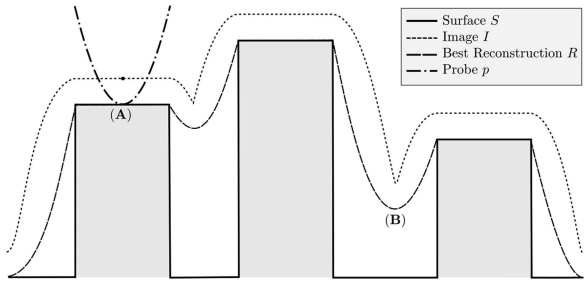


Fig. 3. Graphical depiction of SPM data in 1D. The line depicts an artificial atomic surface $S$, the dotted line is the image $I$ after dilating the true surface with the probing element $p$. The dilation is the natural mathematical model of the equipotential (in STM) movement of the probe across the surface. The striped line depicts the least upperbound on the recoverable surface $R$. Notice especially region *(A)* the image $I$ and surface $S$ have the same shape at the maxima disregarding some offset in $y$; and *(B)* a blunt probe cannot fully recover surfaces within crevices between atoms.

Specifically for AFM, contact between probe and surface can bring about wear of the material during data collection. As an example, double apex forming [35] –i.e., loss of probe convexity– may happen at any time. Consequently, the quality of the probe geometry has to be monitored constantly. Estimating probe and surface abrasion is ill-defined because it may happen to either or both structures [36]: the image signal results from both the probe and the surface. Estimating abrasion effects using MLPs and CNNs has previously been studied in [35], [36], [37], [38], [39], [40]. While results are promising, directly using the morphological nature of the problem in its solution is likely to be beneficial.

For the proof-of-principle of the proposed method, SPM image-surface pairs are required. These are not trivially generated and no public dataset exists that provides them. For example, the authors of [41] provide tools to generate data, but make use of an idealized spherical probe implicit in their atomic representation. [40] provides a binary classification task with negative samples due to a variety of reasons: sample drift, no probe contact, scanning problems, etc. These artefacts are not described using mathematical morphology.

### B. Implementation

All proposed methods, notably Morphological Back-propagation and Probe Learning, are implemented using PyTorch [26]. To create the dataset, three primitives are chosen to replicate three distinct probe shapes: a parabolic probe, a pyramid probe, and a parabolic probe with double apex. Especially the third primitive is relevant for practical applications since non-convex probes may negatively affect the quality and validity of AFM measurements. For each primitive, 8 datasets ($N_{train}$=1000, $N_{test}$=1000) of synthetic 2D train are generated with randomized probe geometry. Each sample consists of a scanned image $I$, an artificial surface $S$, and a best reconstruction $R$ which is the erosion of the image $I$ with known probe $p$. This theoretically best reconstruction $R$ is the least upper bound surface that can be recovered from $I$ taking into account the non-invertibility of the measurement. As a result, all evaluations are done against the best reconstruction $R$ since it cannot reasonably be expected that a network predicts a surface better than $R$. If it would, it could hallucinate erroneous details of the true atomic surface $S$. Training is performed on both $I \to R$ and $I \to S$. See Fig. 4 for an example of the 2.5D data, along with a cross-section

TABLE I
PERFORMANCE OF LINEAR AND MORPHOLOGICAL NETWORKS ON SYNTHETIC 2D SPM DATASETS

| | #Params | #Iters | Train $I \rightarrow R$, Test $R$ | | | Train $I \rightarrow S$, Test $R$ | | |
|---|---|---|---|---|---|---|---|---|
| | | | parabolic | pyramid | double | parabolic | pyramid | double |
| Linear Networks | | | | | | | | |
| 1-layer | 289 | 1e3 | .2525±.1656 | .3904±.2383 | .3973±.1352 | .3744±.1707 | .6833±.1561 | .4512±.1090 |
| 3-layer | 296K | 1e3 | .1908±.0178 | .2421±.0338 | .3109±.1579 | .2967±.0409 | .5802±.1252 | .3817±.0434 |
| U-Net [43] | 13M | 1e4 | .0803±.0051 | .1185±.0049 | .1166±.0077 | .3042±.0205 | .5990±.0514 | .3778±.0397 |
| ResNet-50 [42] | 44M | 1e4 | .1120±.0055 | .1544±.0074 | .1471±.0076 | .2950±.0199 | .5871±.0483 | .3690±.0359 |
| Morphological Networks | | | | | | | | |
| Mean Back-prop | 289 | 1e5 | .0030±.0005 | .0016±.0001 | .0024±.0002 | .2351±.0149 | .4832±.0372 | .2896±.0283 |
| Morph. Back-prop | 289 | 1e3 | **.0000±.0000** | **.0000±.0000** | **.0000±.0000** | .6883±.1254 | .6132±.1327 | 1.0241±.1050 |
| Probe Learning | 289 | 1e2 | .0012±.0007 | .0021±.0006 | .0023±.0008 | **.0009±.0006** | **.0022±.0012** | **.0117±.0014** |

In the second column, the number of parameters is shown; e.g., ResNet-50 [42] has roughly 44m parameters, whereas the simple networks have only 289. In the third column, the average magnitude of iterations is shown until the model converged. In the fourth through ninth column RMSE is shown for predictions measured against the best reconstruction $R$. The networks were independently trained and tested 8 times on different datasets for each configuration. In all cases the morphological networks outperform linear networks. Probe learning performs best, shown in boldface.

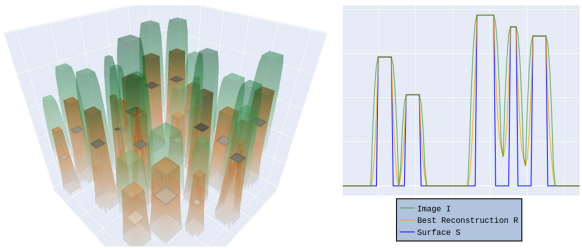


Fig. 4. Data example. *(left)* 2D sample used for training and testing generated by a parabolic probe. Vertical offset added for visualization; *(right)* cross-section along a scanline to show image $I$, surface $S$, and reconstruction $R$. Either $S$ or $R$ may be used as ground truth.

along an arbitrary scanline. The structuring elements of the MNNs are initialized at zero. Initial experimentation shows no impact on performance for random initialization, although convergence time may be slightly affected. For further details, see the publicly available code at github.com/rickgroen/probe-learning.

### C. Learning Probe Geometry

Morphological Back-propagation and Probe Learning by error bounding are evaluated against linear methods and mean back-propagation in Table I. Qualitative examples are shown in Fig. 5. There are four aspects to take note of: first, the single-layer morphological networks outperform, by a large margin, the linear methods. Probe Learning performs best in terms of prediction quality (Fig. 5(a)). For linear methods, high frequency noise is hallucinated around object edges. This effect is more pronounced for U-Net and ResNet (Fig. 5(d)) than for a single-layer CNN (Fig. 5(c)). Second, convolution can be made to perform better by using more parameters and introducing additional tricks such as residual connections. Even so, increasing the amount of parameters in the network to 44 million (shown in the second column in Table I) does not guarantee learning the data; by contrast, the single-layer morphological networks can learn the task by just 289 parameters from a 17×17 probe. Third, the third column in Table I shows the average magnitude until learning converges. Early stopping after RMSE convergence over 100 iterations on the training set is applied. Probe Learning appears to converge always within 100 iterations of training. Fourth, morphological back-propagation fits the data perfectly when trained on $I \rightarrow R$, but provides wrong vertical scaling when trained on $I \rightarrow S$, though overall shape is predicted correctly (Fig. 5(b)). A second MM layer could compensate for this

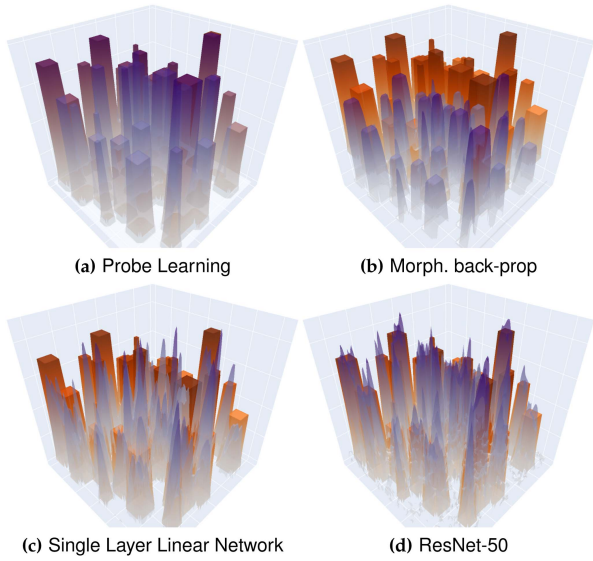


Fig. 5. *(a)* Predictions from a single-layer MNN that learned probe geometry by means of the method in Section II-E. The blue prediction lines up with the orange target $S$; predictions are draped around the true peaks since the network estimates at most within a theoretical upper bound of reconstructability. *(b)* Morphological back-propagation from Section II-C. *(c)* Single-layer CNN; *(d)* ResNet-50. The two CNNs hallucinate erroneous high frequency details around peaks which is more pronounced for *(d)*.

vertical offset or a bias term could be used; alternatively, we could change the error aggregation from (13) to averages (Section II-D) making the method less sensitive to extremes.

### D. Double Apex Detection

Probe or material abrasion is a challenging issue in obtaining high-quality scans [35], [36], [37], [40], [44] in AFM. In Table I, in the sixth and ninth column it is shown that Probe Learning by bounding recovers the best reconstruction $R$. Besides surface predictions, the geometric properties of the probe are learned by the morphological layer. See Fig. 6. The proposed method of probe learning recovers the upper bound of the shape of the probe, within some margin of uncertainty between the two peaks. Numerical analysis of the probe could be integrated with measuring software to determine double-apex forming without the need for complex CNN architectures: a 300-parameter MNN suffices.

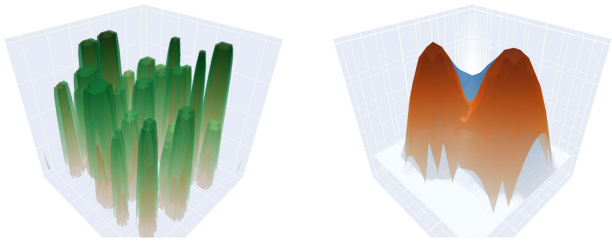

Fig. 6. Qualitative example of double-apex forming in AFM. *(left)* Atomic surface $S$ in orange, resulting image $I$ in green. Artefacts arise as false double peaks around the structures in $I$. Measuring software could mistake these for surface, whereas they actually arise from a low-quality probe. *(right)* The probe used to create the synthetic data in orange, the predicted reconstruction of the shape in blue. The reconstructed probe is an upper bound on the true geometry of the probe. The network thus learns to predict (upper bound) probe geometry.

TABLE II
DEPTH INFILLING PERFORMANCE OF CNNS AND MNNS ON NYU

| | #Params | RMSE ↓ | $\delta < 1.25$ ↑ |
|---|---|---|---|
| **(i)** CNN | 492 | $1.4966 \pm 0.0883$ | $0.2062 \pm 0.0196$ |
| **(ii)** CNN | 46.5K | $1.2974 \pm 0.0481$ | $0.2657 \pm 0.0182$ |
| **(i)** MNN (Mean Back-prop) | 492 | $\mathbf{0.2084 \pm 0.0015}$ | $0.9743 \pm 0.0012$ |
| **(ii)** MNN (Morph. Back-prop) | 492 | $0.2249 \pm 0.0011$ | $\mathbf{0.9771 \pm 0.0001}$ |
| **(iii)** MNN (Probe Learning) | 492 | $0.2241 \pm 0.0017$ | $0.9686 \pm 0.0007$ |

Despite a large number of parameters, CNNs can poorly impute large regions of zero-valued data. In MNNs, the missing values can initially be set to minus or plus infinity, and be processed accordingly by the network. Results are averaged over 10 runs for each model.

## IV. OBSERVATIONS ON GENERALIZATION

Back-propagation naturally generalizes to networks with an arbitrary number of layers. The simulated SPM experiments only address the method on idealized noise-free haptic data; noise impacts network learning. To demonstrate the proposed method, it is applied to depth infilling, i.e., augmenting incomplete depth data. Though specialized algorithms exist [45], infilling is fundamentally a morphological task since it is one of separating shapes and overcoming occlusion.

The depth infilling experiment is performed on NYUv2 ($N_{train}$=795, $N_{test}$=654) [46]. All networks consist of six layers: the first three layers down-sample to $\frac{1}{8}^{\text{th}}$ resolution, the latter three layers up-sample to full resolution. MNNs use dilation and erosion layers alternately, CNNs use convolutions without non-linearities. Networks are trained for 40 epochs using an SGD optimizer and an L2 loss objective. Quantitative results are shown in Table II, visual results are shown in Fig. 7. Since the data are gathered from real-world indoor scenes, it can be expected that sensor noise complicates learning. While morphological networks may be suited to dealing with missing data, noisy data is challenging: morphological operations deal with additive noise by estimating an envelope. For a sequence of morphological layers, without aggregation, the envelope is slightly vertically displaced with respect to the signal. To compensate for this non-morphological type of noise, the morphological layers have to be extended by a vertical bias term; convolutional layers use an identical term to deal with vertical off-set.

## V. CONCLUSION

In this paper, a geometric definition of Morphological Back-propagation is proposed that does not rely on linear approximation of morphological operations but rather on the geometric provenance of slope correspondences. Second, Morphological Probe Learning is proposed based on the natural bounding properties of morphology.

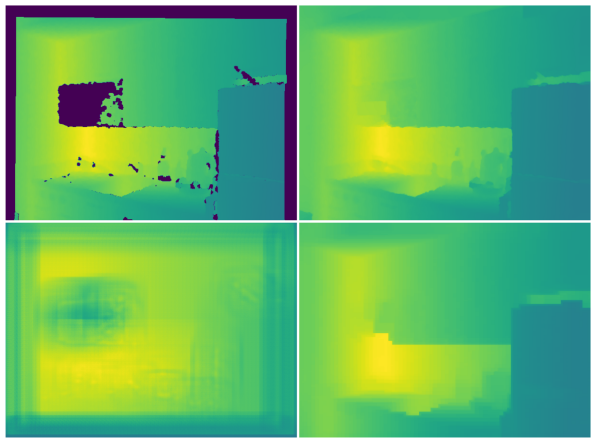

Fig. 7. *(left-upper)* input raw depth; *(right-upper)* ground truth infilled depth; *(left-lower)* CNN *(ii)* prediction; *(right-lower)* MNN *(ii)* prediction. MNNs can deal with sparse data. The size of the structuring element determines how much missing data is filled for each layer.

Two experiments (SPM surface reconstruction and NYU depth infilling) confirm that problems of a morphological nature can be solved accurately with much smaller MMNs than CNNs, and compete with dedicated solutions. In both experiments, CNNs are not able to approximate the ground truth even when many more parameters were introduced. In the case of SPM data, MNNs also converged orders of magnitude faster than their linear counterparts.

As of now, only relatively simple single-channel networks were examined. In future research, the proposed update rules (morphological back-propagation and probe learning) or combinations of both could feature more prominently in larger morphological networks. They should then take into account the composition of the morphological operations in subsequent layers. Moreover, the morphological update rules should be adapted to be less sensitive to noise. In conclusion, when data can reasonably be modelled to result from probing touch (e.g., haptic data from SPM or LiDAR), morphological operations are strongly recommended in the construction of network architectures.

## ACKNOWLEDGMENT

We sincerely thank the anonymous reviewers and our colleague Rein van den Boomgaard for their helpful comments on earlier versions of this manuscript.

## REFERENCES

[1] J. Serra, *Image Analysis and Mathematical Morphology*. New York, NY, USA: Academic, 1983.
[2] J. Serra and L. Vincent, "An overview of morphological filtering," *Circuits Syst. Signal Process.*, vol. 11, no. 1, pp. 47–108, 1992.
[3] L. Dorst and R. van den Boomgaard, "Morphological signal processing and the slope transform," *Signal Process.*, vol. 38, no. 1, pp. 79–98, 1994.
[4] L. Dorst and R. van den Boomgaard, "The systems theory of contact," in *Proc. Int. Workshop Algebr. Frames Perception-Action Cycle*, Springer Verlag, 2000, pp. 22–47.
[5] H. J. Heijmans and R. van den Boomgaard, "Algebraic framework for linear and morphological scale-spaces," *J. Vis. Commun. Image Representation*, vol. 13, no. 1/2, pp. 269–301, 2002.
[6] G. X. Ritter and P. Sussner, "An introduction to morphological neural networks," in *Proc. 13th Int. Conf. Pattern Recognit.*, 1996, pp. 709–717.
[7] P. Sussner, "Morphological perceptron learning," in *Proc. IEEE Int. Symp. Intell. Control Held Jointly IEEE Int. Symp. Comput. Intell. Robot. Automat. Intell*, 1998, pp. 477–482.

[8] V. Charisopoulos and P. Maragos, "Morphological perceptrons: Geometry and training algorithms," in *Proc. Int. Symp. Math. Morphol. Appl. Signal Image Process.*, Springer, 2017, pp. 3–15.

[9] Y. Shen, X. Zhong, and F. Y. Shih, "Deep morphological neural networks," 2019, *arXiv: 1909.01532*.

[10] N. Dimitriadis and P. Maragos, "Advances in the training, pruning and enforcement of shape constraints of morphological neural networks using tropical algebra," 2020, *arXiv: 2011.07643*.

[11] R. Mondal, P. Purkait, S. Santra, and B. Chanda, "Morphological networks for image de-raining," in *Proc. Int. Conf. Discrete Geometry Comput. Imagery*, Springer, 2019, pp. 262–275.

[12] R. Mondal, M. S. Dey, and B. Chanda, "Image restoration by learning morphological opening-closing network," *Math. Morphol.-Theory Appl.*, vol. 4, no. 1, pp. 87–107, 2020.

[13] G. Franchi, A. Fehri, and A. Yao, "Deep morphological networks," *Pattern Recognit.*, vol. 102, 2020, Art. no. 107246.

[14] K. Nogueira, J. Chanussot, M. D. Mura, W. R. Schwartz, and J. A. D. Santos, "An introduction to deep morphological networks," 2019, *arXiv: 1906.01751*.

[15] M. Sangalli, S. Blusseau, S. Velasco-Forero, and J. Angulo, "Scale equivariant neural networks with morphological scale-spaces," in *Proc. Int. Conf. Discrete Geometry Math. Morphol.*, Springer, 2021, pp. 483–495.

[16] D. Worrall and M. Welling, "Deep scale-spaces: Equivariance over scale," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2019, pp. 7364–7376.

[17] H. ElNaghy and L. Dorst, "Boundary morphology for hierarchical simplification of archaeological fragments," *Math. Morphol.-Theory Appl.*, vol. 4, no. 1, pp. 46–63, 2020.

[18] S. Velasco-Forero, R. Pagès, and J. Angulo, "Learnable empirical mode decomposition based on mathematical morphology," *SIAM J. Imag. Sci.*, vol. 15, no. 1, pp. 23–44, 2022.

[19] M. Nakashizuka, S. Takenaka, and Y. Iiguni, "Learning of structuring elements for morphological image model with a sparsity prior," in *Proc. IEEE Int. Conf. Image Process.*, 2010, pp. 85–88.

[20] J. Masci, J. Angulo, and J. Schmidhuber, "A learning framework for morphological operators using counter–harmonic mean," in *Proc. Int. Symp. Math. Morphol. Appl. Signal Image Process.*, Springer, 2013, pp. 329–340.

[21] P. Salembier, "Structuring element adaptation for morphological filters," *J. Vis. Commun. Image Representation*, vol. 3, no. 2, pp. 115–136, 1992.

[22] Y. Hu, N. Belkhir, J. Angulo, A. Yao, and G. Franchi, "Learning deep morphological networks with neural architecture search," 2021, *arXiv:2106.07714*.

[23] R. Mondal, S. Santra, S. S. Mukherjee, and B. Chanda, "Morphological network: How far can we go with morphological neurons?," in *Proc. 33rd Brit. Mach. Vis. Conf.*, BMVA Press, 2022, Art. no. 779.

[24] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[25] R. van den Boomgaard and A. Smeulders, "The morphological structure of images: The differential equations of morphological scale-space," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 11, pp. 1101–1113, Nov. 1994.

[26] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2019, pp. 8024–8035.

[27] J. S. Villarrubia, "Morphological estimation of tip geometry for scanned probe microscopy," *Surf. Sci.*, vol. 321, no. 3, pp. 287–300, 1994.

[28] N. Bonnet, S. Dongmo, P. Vautrot, and M. Troyon, "A mathematical morphology approach to image formation and image restoration in scanning tunnelling and atomic force microscopies," *Microsc. Microanalysis Microstructures*, vol. 5, no. 4–6, pp. 477–487, 1994.

[29] D. L. Wilson, K. S. Kump, S. J. Eppell, and R. E. Marchant, "Morphological restoration of atomic force microscopy images," *Langmuir*, vol. 11, no. 1, pp. 265–272, 1995.

[30] L. Dongmo, J. S. Villarrubia, S. N. Jones, T. B. Renegar, M. T. Postek, and J.-F. Song, "Experimental test of blind tip reconstruction for scanning probe microscopy," *Ultramicroscopy*, vol. 85, no. 3, pp. 141–153, 2000.

[31] J. S. Villarrubia, "Algorithms for scanned probe microscope image simulation, surface reconstruction, and tip estimation," *J. Res. Nat. Inst. Standards Technol.*, vol. 102, no. 4, 1997, Art. no. 425.

[32] G. Sacha, F. Rodriguez, and P. Varona, "An inverse problem solution for undetermined electrostatic force microscopy setups using neural networks," *Nanotechnology*, vol. 20, no. 8, 2009, Art. no. 085702.

[33] I. Azuri, I. Rosenhek-Goldian, N. Regev-Rudzki, G. Fantner, and S. R. Cohen, "The role of convolutional neural networks in scanning probe microscopy: A review," *Beilstein J. Nanotechnol.*, vol. 12, no. 1, pp. 878–901, 2021.

[34] S. V. Kalinin et al., "Big, deep, and smart data in scanning probe microscopy," *ACS Nano*, vol. 10, no. 10, pp. 9068–9086, 2016.

[35] M. Rashidi and R. A. Wolkow, "Autonomous scanning probe microscopy in situ tip conditioning through machine learning," *ACS Nano*, vol. 12, no. 6, pp. 5185–5189, 2018.

[36] P. Bakucz, A. Yacoot, T. Dziomba, L. Koenders, and R. Krüger-Sehm, "Neural network approximation of tip-abrasion effects in AFM imaging," *Meas. Sci. Technol.*, vol. 19, no. 6, 2008, Art. no. 065101.

[37] B. W. Erickson, S. Coquoz, J. D. Adams, D. J. Burns, and G. E. Fantner, "Large-scale analysis of high-speed atomic force microscopy data sets using adaptive image processing," *Beilstein J. Nanotechnol.*, vol. 3, no. 1, pp. 747–758, 2012.

[38] A. A. Vekinis and V. Constantoudis, "Neural network evaluation of geometric tip-sample effects in AFM measurements," *Micro Nano Eng.*, vol. 8, 2020, Art. no. 100057.

[39] O. M. Gordon, F. L. Junqueira, and P. J. Moriarty, "Embedding human heuristics in machine-learning-enabled probe microscopy," *Mach. Learn.: Sci. Technol.*, vol. 1, no. 1, 2020, Art. no. 015001.

[40] A. Krull, P. Hirsch, C. Rother, A. Schiffrin, and C. Krull, "Artificial-intelligence-driven scanning probe microscopy," *Commun. Phys.*, vol. 3, no. 1, pp. 1–8, 2020.

[41] K. Choudhary et al., "Computational scanning tunneling microscope image database," *Sci. Data*, vol. 8, no. 1, pp. 1–9, 2021.

[42] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[43] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, Springer, 2015, pp. 234–241.

[44] O. Gordon et al., "Scanning tunneling state recognition with multi-class neural network ensembles," *Rev. Sci. Instrum.*, vol. 90, no. 10, 2019, Art. no. 103704.

[45] A. Levin, D. Lischinski, and Y. Weiss, "Colorization using optimization," in *Proc. ACM SIGGRAPH Papers*, 2004, pp. 689–694.

[46] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from RGBD images," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 746–760.