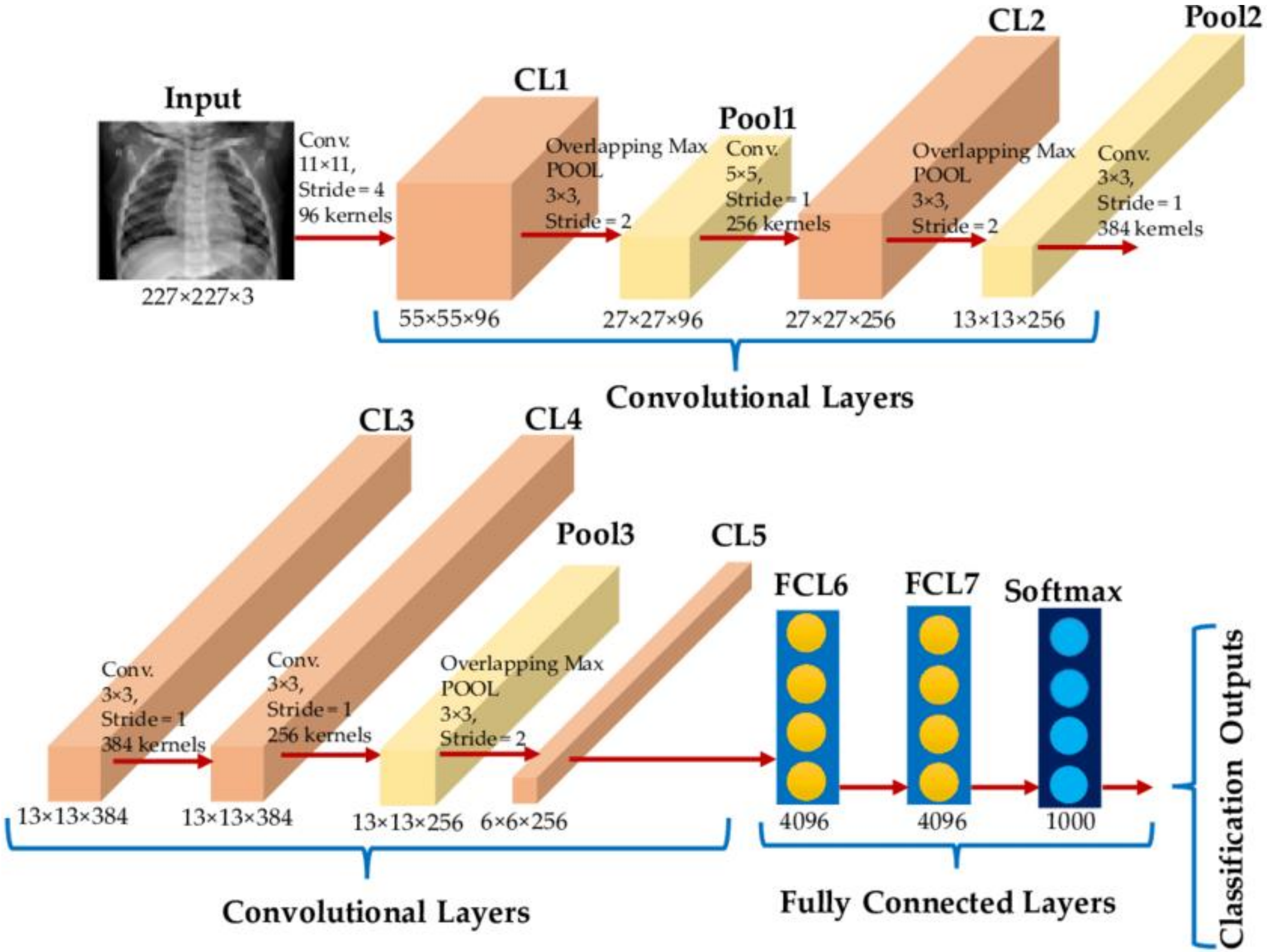
Imaging With Equivariant Deep Learning

From unrolled network design to fully unsupervised learning

CONTENTS

1. Introduction



Symmetry in AlexNet

For a function $f: x \to y$ and a transformation $g: x \to x$

Equivariant:

$$f(g(x)) = g(f(x))$$

Invariant:

$$f(x) = f(g(x))$$

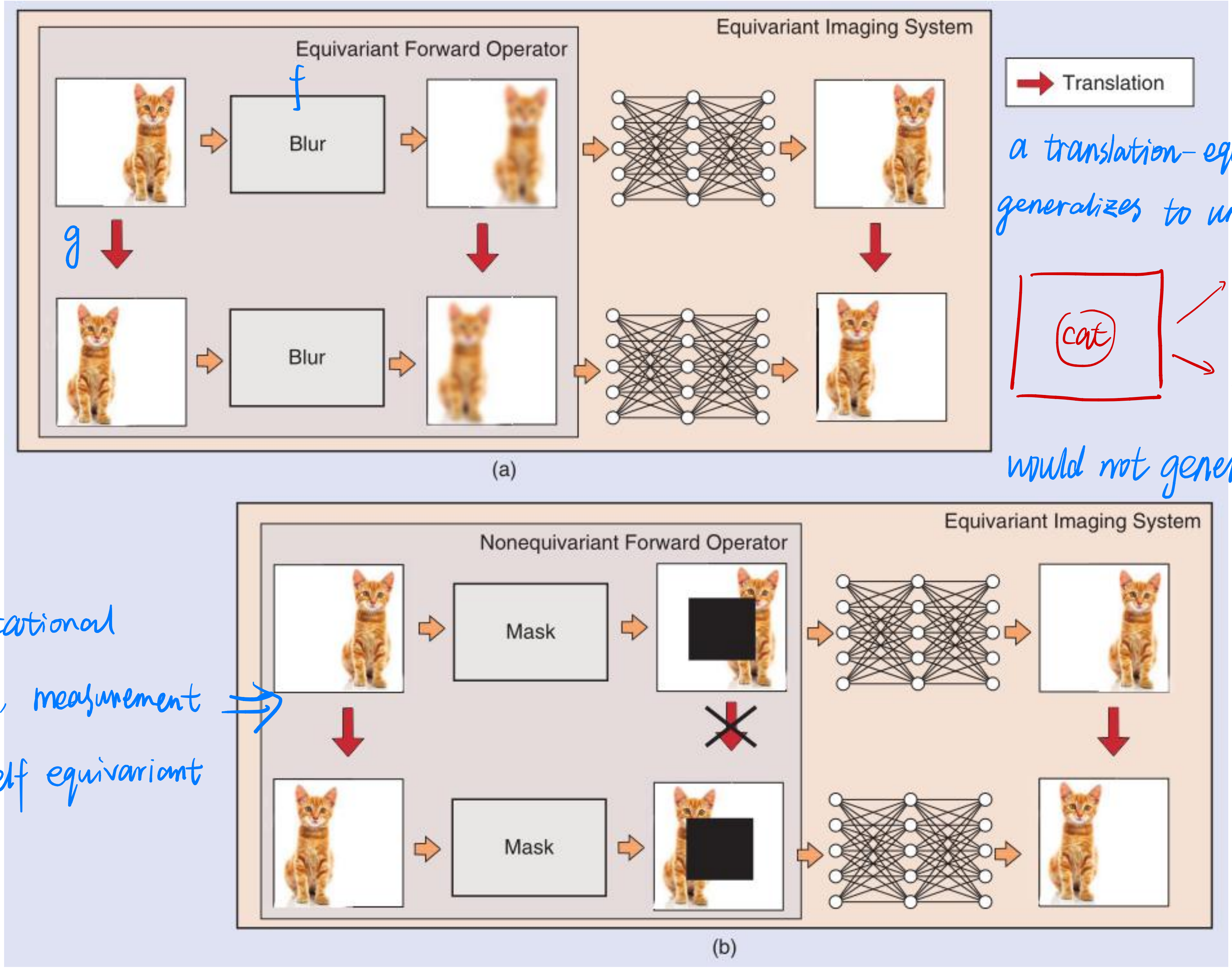
- **Architectural design**

Convolution, max pooling

- **Training process**

Data Augmentation

reconstruction



a translation-equivariant network generalizes to unseen translations.

can reconstruct

cat

cannot

would not generalize to unseen.

In many computational imaging problems, measurement process is not itself equivariant

2. Group actions and equivariance

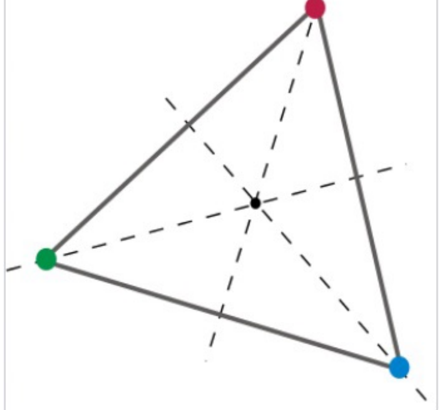
- A **group (G, ·)** is **a set G** equipped with **a product ·**
- Properties:
  - (1) closed: G×G→G
  - (2) associative: $g_1 \cdot (g_2 \cdot g_3) = (g_1 \cdot g_2) \cdot g_3$
  - (3) identity element: exist $e \in$ G, s.t. $g \cdot e = e \cdot g$
  - (4) inverse element: exist $g^{-1} \in$ G, s.t. $g \cdot g^{-1} = g^{-1} \cdot g$
- simply write $g \cdot h = gh$ and refer to the group by the name of the underlying set G
- eg. Z and +

$$g_{p4}(r, u, v) = \begin{bmatrix} \cos\left(\frac{2r}{2}\right) & -\sin\left(\frac{2r}{2}\right) & u \\ \sin\left(\frac{2r}{2}\right) & \cos\left(\frac{2r}{2}\right) & v \\ 0 & 0 & 1 \end{bmatrix}$$

$r, (u, v) \in \{0, 1, 2, 3\}, Z^2$

Group action

The cyclic group C_3 acts on the set of the three vertices through rotations.



The concept of a group is particularly interesting when combined with the concept of an action: given a (potentially abstract) group G and a set X, we will say that G acts on X through T if $T = \{T_g : X \to X\}_{g \in G}$ is a collection of invertible transformations that is compatible with the group, in the sense that

$$T_{g_1} \circ T_{g_2} = T_{g_1 \cdot g_2}$$

T_g can be seen as a operator form of g.

- A particularly simple group action is the trivial action of G on X, in which the case the group "acts" by doing nothing: $T_g = \mathrm{id}_X$ for all $g \in G$.

X : pixel coordinates Y : pixel values

In this work, we are concerned with images, in which case the signals of interest can usually be modeled as functions $u : X \to Y$. As an example, for color images X is a subset of $\mathbb{R}^2$ and Y is $\mathbb{R}^3$. A group action T of G on the domain X can be lifted to an action T' on the set of signals by $T'_g(u)(x) = u(T_{g^{-1}}(x))$. If the set of signals is a vector space (as for the color images), T' acts linearly on the signals, making it a so-called representation of G. Similarly, an action T of G on the range Y of the signals can be lifted to an action on the signals, T', by $T'_g(u)(x) = T_g(u(x))$, and in fact actions on the domain and range can be combined if so desired.

The affine group on a Euclidean space $\mathbb{R}^d$: all transformation of the form $x \mapsto Hx + h$
$H \in \mathbb{R}^{d \times d}$, $h \in \mathbb{R}^d$

FIGURE 2. An illustration of group actions on images. Groups can act on images by transforming either their range or their domain. (a) Reference image. (b) Transformation of range: e.g., color inversion. (c) Transformation of domain: e.g., translation, rotation, scaling, and shearing.

In the setting of computational imaging, we are concerned with maps $\Phi : \mathcal{X} \to \mathcal{Y}$ between potentially different sets $\mathcal{X}$ and $\mathcal{Y}$ representing spaces of images and/or measurements. For example, Φ could be the forward operator or a reconstruction operator. If both $\mathcal{X}$ and $\mathcal{Y}$ share a symmetry in the form of potentially different group actions T of G on $\mathcal{X}$ and T' of G on $\mathcal{Y}$, we may ask whether Φ respects these symmetries, in the following sense:

Equivariance: we call Φ equivariant, if $\Phi(T_g(u)) = T'_g(\Phi(u))$ holds for all $u \in \mathcal{X}$ and $g \in G$.

Invariance: if T' is the trivial action of G on $\mathcal{Y}$ and Φ is equivariant, we will call Φ invariant. In this case, we have $\Phi(T_g(u)) = \Phi(u)$ for all $u \in \mathcal{X}$ and $g \in G$.

$\forall v \in \mathcal{Y}$, $\Psi(T'_g(v)) = T''_g(\Psi(v))$, replace v with $\Phi(u)$ and $\Phi(T_g(u)) = T'_g(\Phi(u))$

$\Psi(T'_g(\Phi(u))) = T''_g(\Psi \circ \Phi(u)) \iff \Psi(\Phi(T_g(u))) = T''_g((\Psi \circ \Phi)(u))$

An additional fact that will be of importance later is that equivariance is preserved under function composition: if G is a group that acts on spaces $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ through T, T', T'' respectively and $\Phi : \mathcal{X} \to \mathcal{Y}$ and $\Psi : \mathcal{Y} \to \mathcal{Z}$ are equivariant, $\Psi \circ \Phi : \mathcal{X} \to \mathcal{Z}$ is equivariant in the sense that $(\Psi \circ \Phi)(T_g(u)) = T''_g((\Psi \circ \Phi)(u))$. transitivity

3. Computational imaging, equivariance, and deep learning

- Computational imaging, distinct from other forms of image processing, relies on the acquisition of sensor measurements that **indirectly** inform about the imaged object.
- Inverse problem:

forward (acquisition) process

$$\underbrace{y}_{\text{measurement}} = A(\underbrace{u}_{\text{image representation}}) + \epsilon.$$

- where $y \in R^{m}$, $u \in R^{n}$

Model-based image reconstruction(MBIR)

$$\underset{u}{\operatorname{argmin}}\, E(u) + J(u).$$

is composed of a data consistency loss term $E(u) = d(A(u), y)$ to capture the role of the acquisition physics in the measurement process, including the noise statistics, along with a regularization function $J(u)$ that incorporates prior knowledge (e.g., sparsity in the wavelet domain) of u and penalizes less plausible solutions.

- In the continuous domain, total variation (TV) prior is naturally invariant to translations, rotations, and reflections.

(non-differentiable)

- Simple gradient descent methods cannot be directly applied.
- Proximal splitting methods:

For example, a non-differentiable regularization function J can be handled through its proximal mapping which takes the following form: given u , find v that minimize loss.

$$\mathrm{prox}_J(u) = \operatorname*{argmin}_v \frac{1}{2}\|u - v\|_2^2 + J(v) \tag{3}$$

- Basic PGD

algorithm proceeds by taking a step in the negative gradient direction of the smooth component of the cost function, followed by a proximal mapping to reduce the non-differentiable cost. Its update equation thus takes the following simple form:

$$u^{(k)} = \mathrm{prox}_{\tau J}\left(u^{(k-1)} - \tau \nabla E(u^{(k-1)})\right) \tag{4}$$

where $\tau > 0$ denotes the step size of the algorithm.

Deep learning for inverse problems

In particular, due to the powerful representation learning properties of DNNs, a range of neural network solutions have recently been proposed for computational imaging (see [18], [19] for detailed surveys). In this setting, the goal is usually to learn a reconstruction function $f_\theta : y \mapsto u$ parameterized by the network weights θ, using N pairs of measurements and ground-truth images $\{(y_i, u_i)\}_{i=1,\ldots,N}$. The networks are typically trained by minimising the empirical risk

$$\min_\theta \sum_{i=1}^{N} \ell(u_i, f_\theta(y_i)) \tag{5}$$

$y \mapsto u \Longrightarrow A^\dagger y \mapsto u$ incorporate the acquisition physics

but require large quantities of training data.

Alternatively, we can design the architecture of f_θ using ideas of MBIR solutions

Consider again the PGD algorithm in (4). A simple modification replaces the proximal map along with the step size, τ, at the kth iteration with a neural network $f_\theta^{(k)}$, such that:

$$u^{(k)} = f_\theta^{(k)}(u^{(k-1)}, \nabla E(u^{(k-1)})). \tag{6}$$

The algorithm is then run for a fixed number of iterations, $k = 1, \ldots, \mathrm{iter}_{\max}$, as illustrated in Figure 3. The learnable weights in individual networks $\{f_\theta^{(k)}\}$ can be either tied or varied from iteration to iteration.

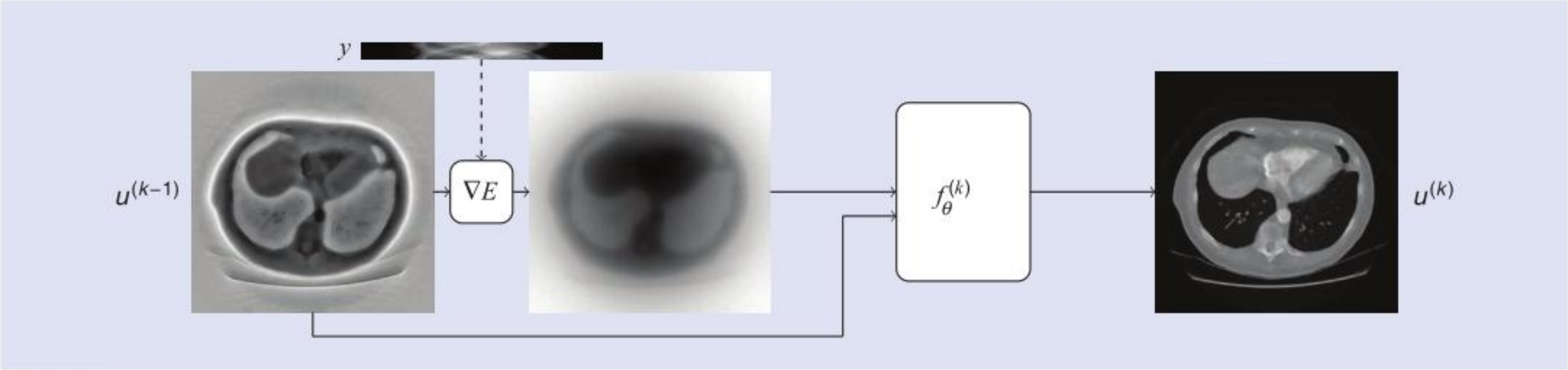


FIGURE 3. A schematic overview of an iteration of an unrolled PGD algorithm applied to the problem of CT reconstruction. In this setting, the inputs $u^{(k-1)}$ and $\nabla E(u^{(k-1)})$ may be combined as in PGD using a (learnable) step size τ to give $u^{(k-1)} - \tau\nabla E(u^{(k-1)})$ before processing with the NN. It is also possible, however, to allow the NN to learn a more general mixing of these inputs.

While such hybrid MBIR-DNN approaches have proved highly successful, providing state-of-the-art imaging solutions, it appears that by adopting these data-driven approaches we may have thrown away the other prior physical knowledge we have, namely the symmetry properties of our signal set. In the next two sections, we review ways to remedy this through either a modified unrolled network architecture or through the training process itself.

4. Equivariance by design

- NNs are alternating compositions of simple linear and nonlinear functions, so we are led to study the problem of designing **linear and nonlinear equivariant functions**.
- For clarity of exposition, this section will treat signals as continuous objects.

One established approach to designing equivariant networks, which leads into the systematic approach that we will study in the next section, can be found in CNNs [24]. Treating an image as a function $u : \mathbb{R}^2 \to \mathbb{R}$, we can act on it with a translation $h \in \mathbb{R}^2$ by $T_h u(x) = u(x - h)$. In this case, convolution by an arbitrary filter $k : \mathbb{R}^2 \to \mathbb{R}$, i.e. $u \mapsto k * u$, is equivariant, where

$k * (T_h u(x)) = T_h(k * u(x))$

$$k * u(x) = \int_{\mathbb{R}^2} k(x')u(x - x')\, \mathrm{d}x'. \tag{7}$$

Equivariant neural networks

Lifting approach: It is possible to generalise the Euclidean convolution of Equation (7) to a group convolution, which combines two signals defined on the group in an equivariant manner. Under a technical condition (local compactness) that is satisfied for many groups, it is possible to define an invariant measure μ (the so-called Haar measure) on the group G. This invariance means that for any integrable $u : G \to \mathbb{R}$ and group element $g \in G$, we have

if $G = \mathbb{R}^d$, Lebesgue

$$\int_G u(gh)\,\mathrm{d}\mu(h) = \int_G u(h)\,\mathrm{d}\mu(h). \tag{8}$$

- With this measure, we can define equivariant convolutions on the group by

$T_h u(g) = u(h^{-1}g)$

$$k * u(g) = \int_G k(h)u(h^{-1}g)\mathrm{d}\mu(h). \tag{9}$$

(7) $G = \mathbb{R}^2$ is the group of translations

Note that this convolution acts on signals that have as domain G. This is where the lifting name comes into play: an input signal such as an image generally has as domain a space such as $\mathbb{R}^d$, i.e. it is not of the form required to apply the Equation (9). To prepare such an ordinary input signal, we need to "lift" it to G, for instance using a linear map such as

$\mathbb{R}^d \to G$

$$Lu(g) = \int_{\mathbb{R}^d} k(g^{-1}x)u(x)\,\mathrm{d}x, \tag{10}$$

where $k : \mathbb{R}^d \to \mathbb{R}$ is again a filter with learnable parameters. This approach was pioneered in [10],

$x \to$ the domain of an image $\mathbb{R}^d$

$g^{-1}x \to g^{-1}$ acts on x (many different forms)

- **Steerable filters approach**

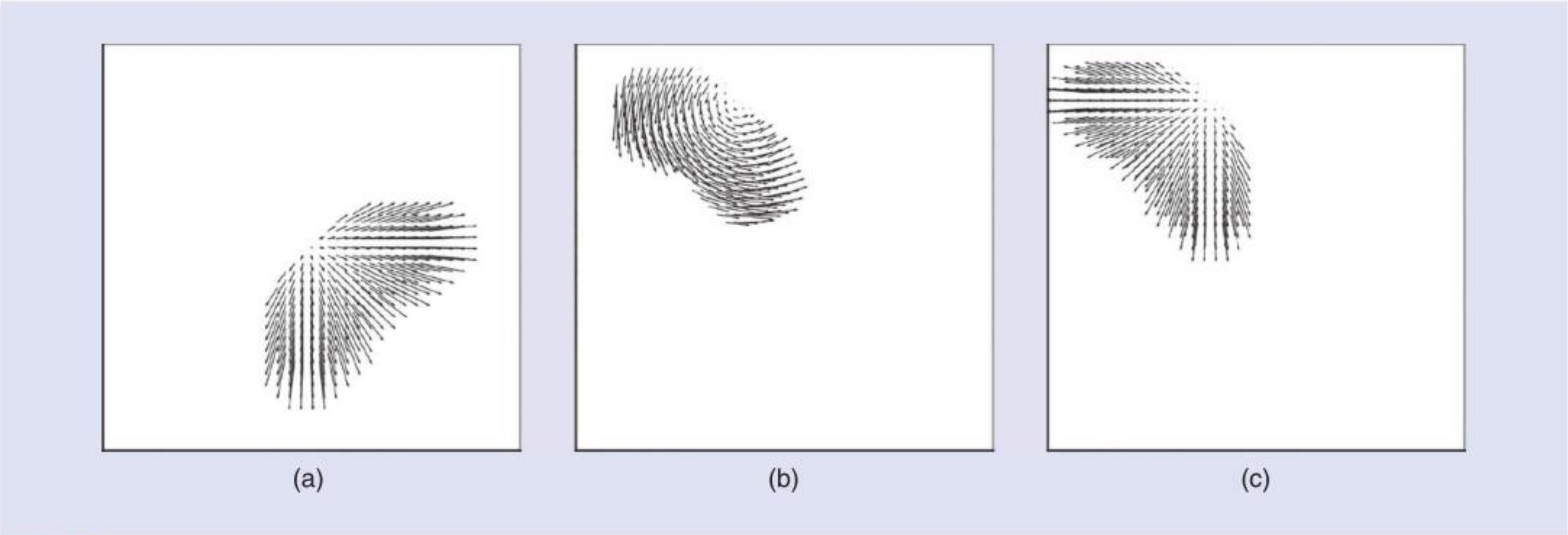


FIGURE 4. To properly transform geometric features, such as the vector field shown here, it is necessary for the group action on the domain to be followed by a group action on the range. In this case, we have a vector field, so the representation π is simply given by $\pi_H = H$. (a) $u(x)$. (b) $u(H^{-1}(x-h))$. (c) $\pi_H u(H^{-1}(x-h))$.

Group contains all translations and that the transformations are isometries.

- For more complicated geometric features that are not just vector fields, this is generalised by transforming the range of the signal using a representation π of the linear transformations being considered. Correspondingly, the group actions we consider will all be of the form

$$T^{\pi}_{(h,H)}u(x) = \pi_H u(H^{-1}(x - h)), \tag{11}$$

where (h, H) is a group element, consisting of a translation $h \in \mathbb{R}^d$ and a linear operator $H : \mathbb{R}^d \to \mathbb{R}^d$, and π_H is a representation of the linear operator H, acting on the range of the signal $u : \mathbb{R}^d \to \mathbb{R}^{d_\pi}$.

We assume that there is such a group action on the input signals and a similarly defined group action $T^{\pi'}$ on the output signals for a potentially different representation π' of the linear operators under consideration. The goal is to design an equivariant convolution mapping input signals $u_{\text{in}} : \mathbb{R}^d \to \mathbb{R}^{d_\pi}$ to output signals $u_{\text{out}} : \mathbb{R}^d \to \mathbb{R}^{d_{\pi'}}$ by

$$k * u(x) = \int_{\mathbb{R}^d} k(x')u(x - x')\,\mathrm{d}x'$$

Writing out the equivariance condition, we find that

$$\pi'_H \int_{\mathbb{R}^d} k(x')u(H^{-1}(x-h)-x')\,\mathrm{d}x' = T^{\pi'}_{(h,H)}[u * k](x)$$

$$\pi'_H (k * u (H^{-1}(x-h)))$$

$$= [T^{\pi}_{(h,H)}u] * k(x) \qquad \text{(Equivariance)}$$

$$= \int_{\mathbb{R}^d} k(x')\pi_H u(H^{-1}(x-h-x'))\,\mathrm{d}x'$$

$$= \int_{\mathbb{R}^d} k(Hx')\pi_H u(H^{-1}(x-h)-x')\,\mathrm{d}x'.$$

(Change of variables, using that H is an isometry)

Rearranging, we have

$$0 = \int_{\mathbb{R}^d} (\pi'_H k(x') - k(Hx')\pi_H)u(H^{-1}(x-h)-x')\,\mathrm{d}x'$$

and since this must hold for arbitrary signals u, we find that this is equivalent to asking that the kernel $k : \mathbb{R}^d \to \mathbb{R}^{d_{\pi'} \times d_\pi}$ satisfies the condition

$$k(Hx)\pi_H = \pi'_H k(x) \tag{12}$$

for all linear operators $H : \mathbb{R}^d \to \mathbb{R}^d$ that occur in the group G [26]. This constraint (which is linear in k) can be solved ahead of time, and discretized to give a basis of equivariant convolution kernels.

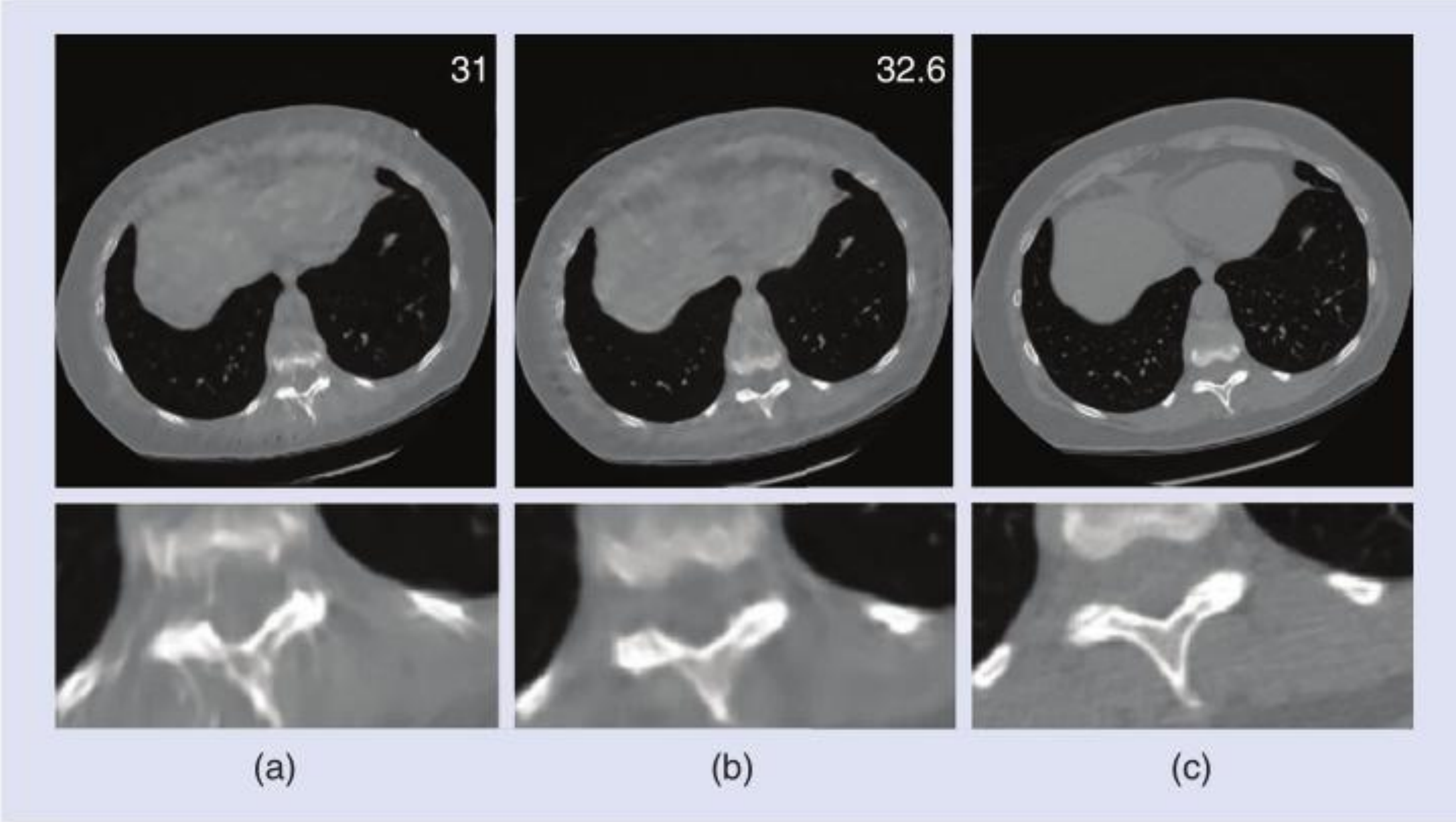
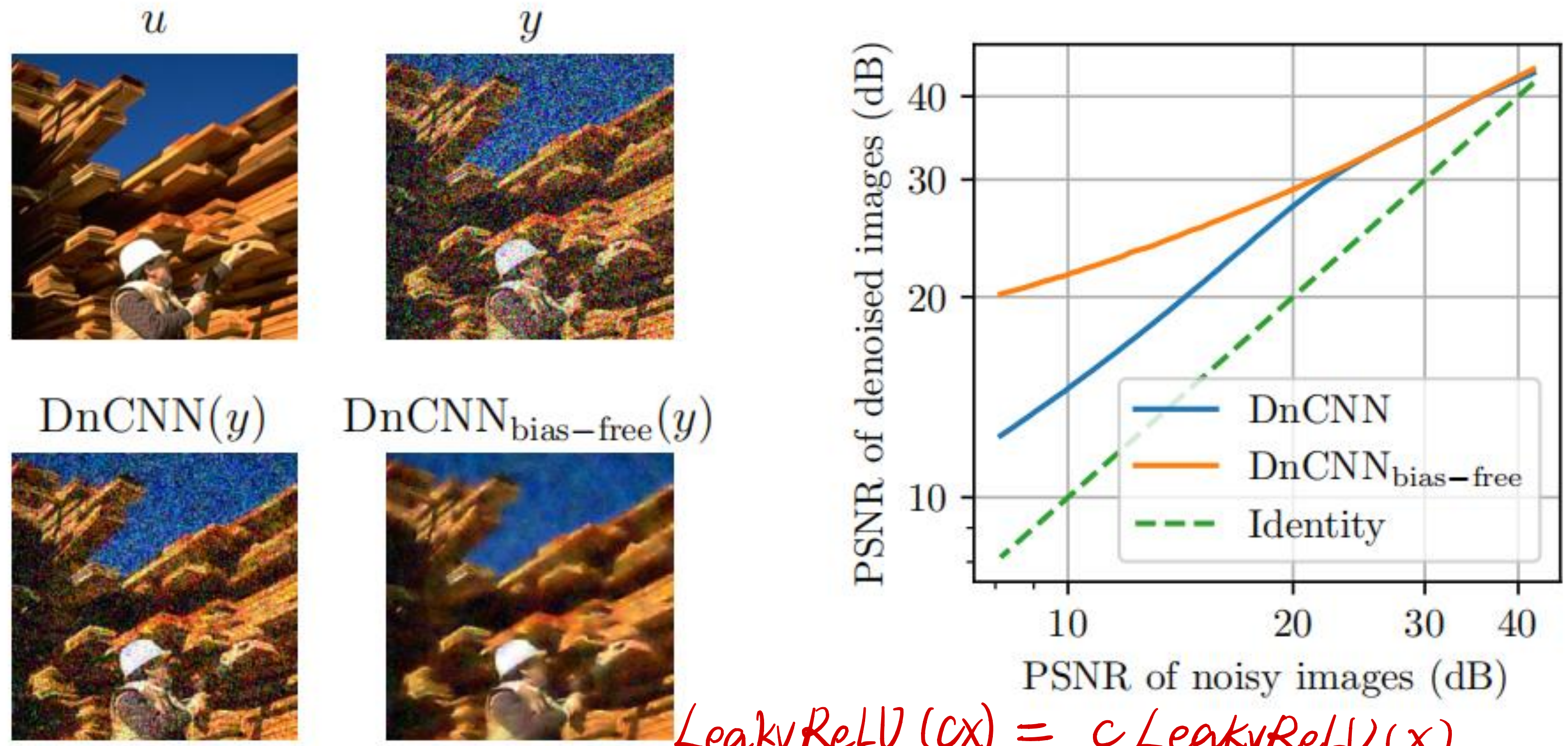


FIGURE 5. A reconstruction with provable equivariant NNs: translational versus rototranslational. Incorporating more inductive bias into the NN improves the PSNR. Even more importantly, the additional inductive bias leads to a better reconstruction of fine details. (a) Translational equivariance, (b) rototranslational equivariance, and (c) reference. PSNR: peak signal-to-noise ratio. (Source: [11].)



$\text{LeakyReLU}(cx) = c\,\text{LeakyReLU}(x)$

$\max\{ax, x\}$, $0<a<1$, scalar multiplication equivariant

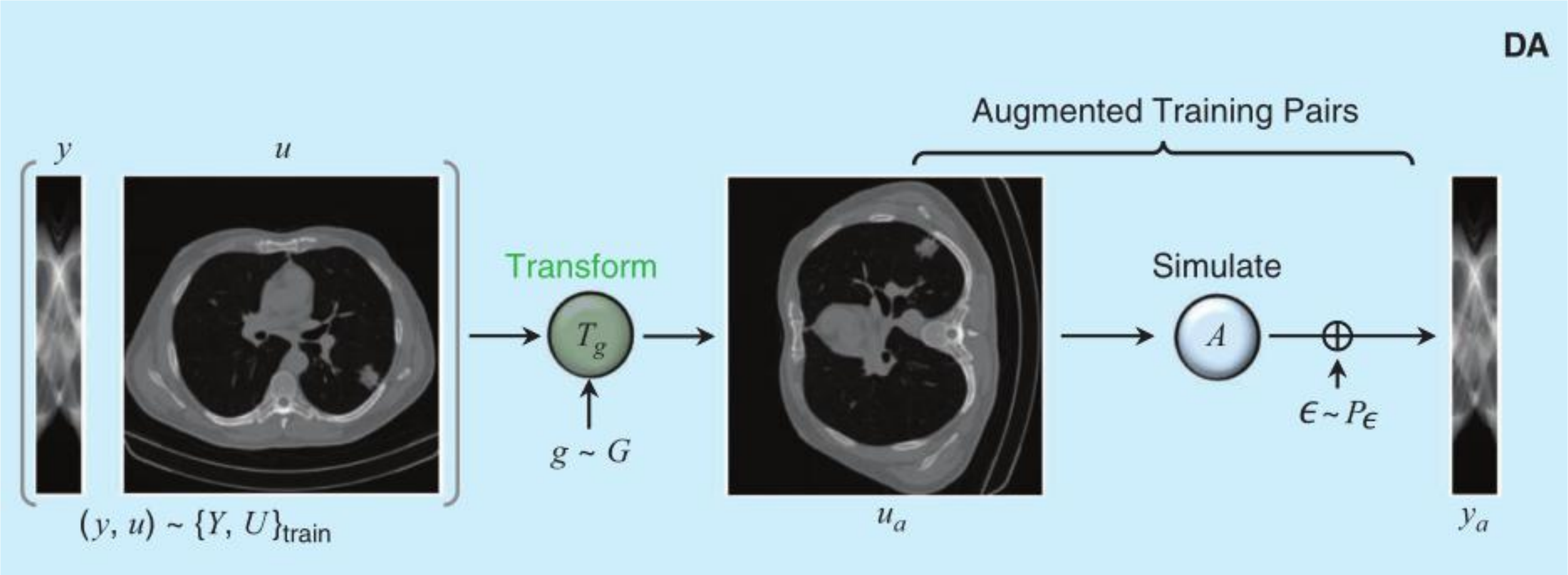
Fig. 6: A comparison of DNN denoisers trained to denoise images corrupted by additive Gaussian white noise, using `LeakyReLU` as activation function. The denoisers are trained on pairs of clean and noisy images with a limited range of noise levels (PSNR $\sim 26-34$ dB) and then tested on a wide range of noise levels. Evidently, the denoiser that does not use biases (which is equivariant to scaling of the range) is vastly more robust to unseen noise levels than the denoiser that does use biases. The noisy image y in this example has a PSNR of 9.9 dB, $\text{DnCNN}(y)$ has a PSNR of 15.2 dB and $\text{DnCNN}_{\text{bias-free}}(y)$ has a PSNR of 21.0 dB.

5. Equivariance by learning

- An alternative way to impose equivariance is to enforce it through the training process instead of using equivariant architectures.
- In the **supervised** setting, where ground-truth images are available, this can be done through **DA** (Data Augmentation)
- while for **unsupervised** learning, **a system-equivariant self-supervised loss** can be used.

Equivariance through DA

- DA is based on the assumption that there is often additional information within the training data that has so far been unused.
- DA introduce a **set of transforms** through which one can modify the existing training data to **generate new plausible samples.**



Equivalence in unsupervised learning

- consider a naive unsupervised loss, which only enforces measurement consistency, e.g.,

$$\sum_{i=1}^{N} \|A f_\theta(y_i) - y_i\|_2^2. \tag{15}$$

There are infinitely many solutions f_θ that attain zero training error.

Perhaps surprisingly though, the weak assumption of invariance to actions of compact groups can be enough for fully unsupervised learning [13]. To understand this, note that such invariance means an observation y can be equally thought of as an observation of a different signal, $\tilde{x}$, via a *virtual* measurement operator $A_g = AT_g$ such that:

$$y = Ax = AT_g T_g^{-1} x = A_g \tilde{x} \tag{16}$$

where group invariance ensures that $\tilde{x} = T_g^{-1} x = T_g^T x$ is a valid element of our signal model. The group action here *rotates* the nullspace of A:

$\mathcal{N}_A = \{ x \mid Ax = 0 \}$

$$\mathcal{N}_{A_g} = T_g^T \mathcal{N}_A \tag{17}$$

$\mathcal{N}_{A_g} = \{ \tilde{x} \mid A_g \tilde{x} = 0 \}$

potentially exposing parts of the original nullspace to view. In order to see the whole of the signal space it is therefore necessary that the concatenation of all the virtual measurement operators,

the group must be big enough

$$M = \begin{bmatrix} AT_1 \\ \vdots \\ AT_{|G|} \end{bmatrix} \tag{18}$$

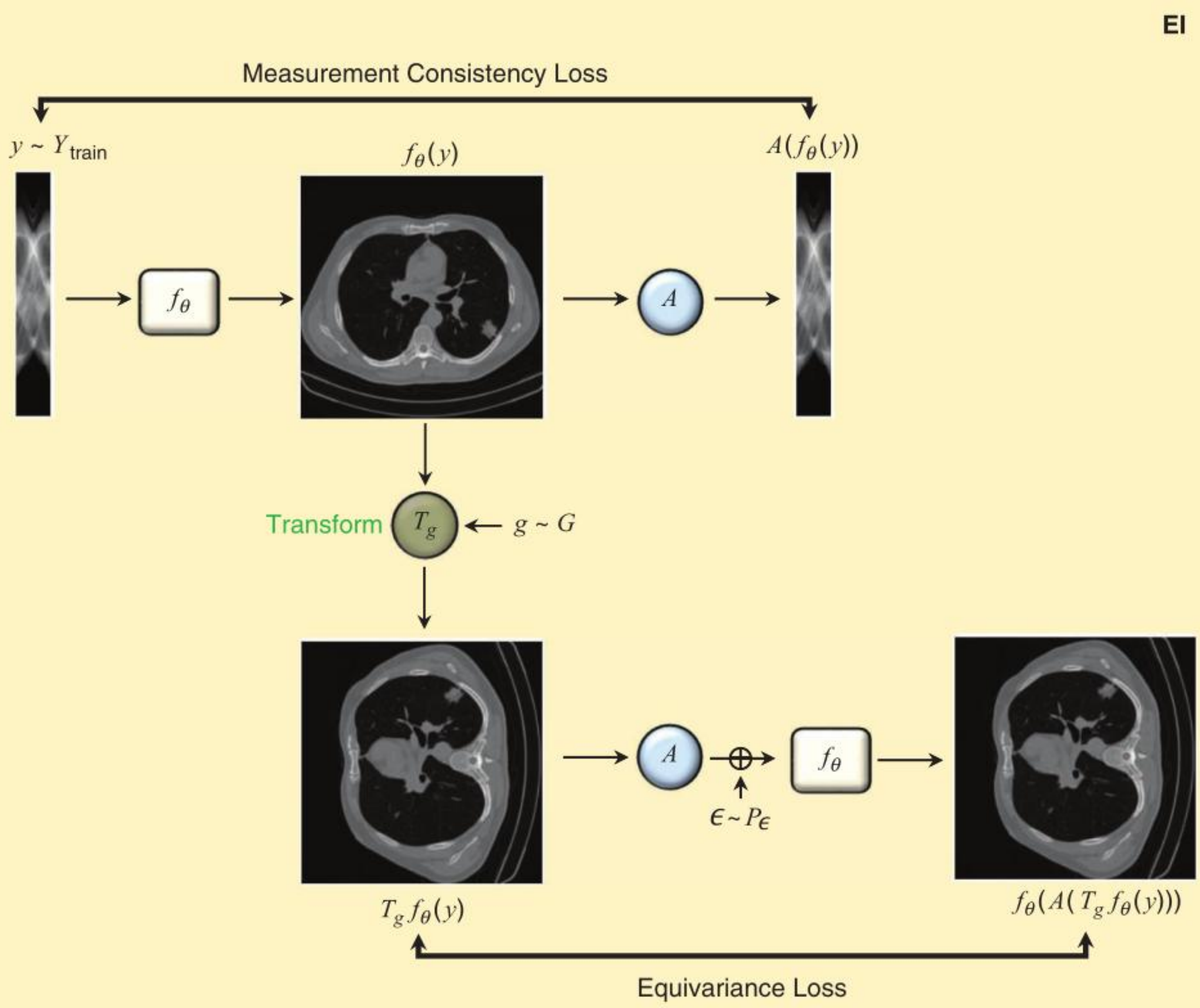
be full rank. One can also think of M here as being the combined measurement operator associated with having oracle simultaneous access to all the virtual measurements of the same signal, x.

EI

1) Equivariant Imaging: If the forward operator is *not* equivariant and the group is big enough, then we can expect to be able to learn from only measurements (an in-depth analysis of the necessary and sufficient conditions for unsupervised learning can be found in [32]). The equivariant imaging (EI) framework [13] offers an elegant way of pursuing system equivariance through self-supervised learning, by using following surrogate loss function:

$$\underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{N} \sum_{g \in G} \|A f_\theta(y_i) - y_i\|_2^2 + \alpha \| f_\theta(A T_g f_\theta(y_i)) - T_g f_\theta(y_i)\|_2^2, \tag{19}$$

- where the first term enforces data consistency , the second term enforces system equivariance, and a controls the strength of equivariance loss.


EI
Measurement Consistency Loss
$y \sim Y_{\text{train}}$
$f_\theta(y)$
$A(f_\theta(y))$
f_θ
A
Transform
T_g
$g \sim G$
A
f_θ
$\epsilon \sim P_\epsilon$
$T_g f_\theta(y)$
$f_\theta(A(T_g f_\theta(y)))$
Equivariance Loss

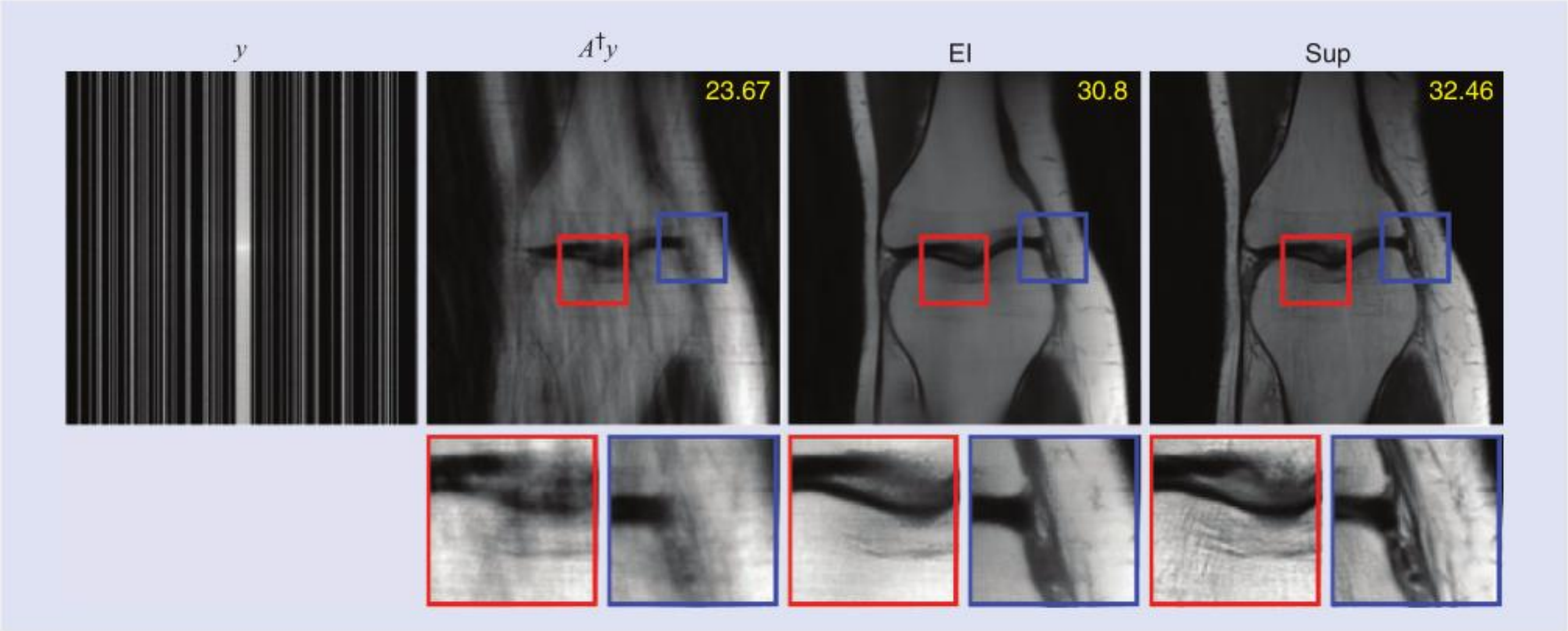


FIGURE 8. A comparison of EI [13] reconstruction with linear inversion and supervised learning for a $4\times$ accelerated single-coil MRI. The network architecture in each case was a U-net taking $A^{\dagger}y$ as the input. By enforcing system equivariance during unsupervised training, EI can perform almost as well as a fully supervised network. PSNR values are shown in the top right corner of the images. Sup: supervised.

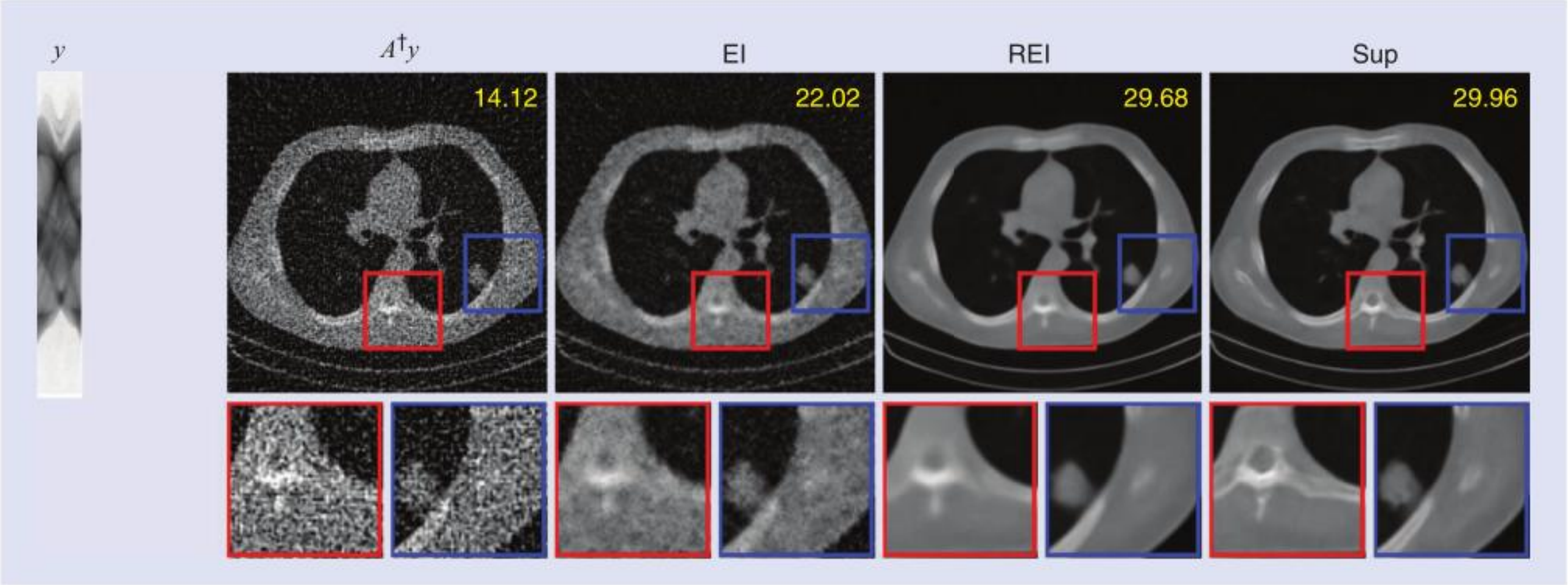


FIGURE 9. Low-dose CT image reconstruction on the test observations (50-view sinograms) with mixed Poisson–Gaussian noise. A comparison among linear inversion, EI, REI, and a supervised learning solution. PSNR values are shown in the top right corners. (Source: [33].)

6. Open problems and future directions

- Opportunities and limitations of equivariance
  - current implementations tend to impose only limited symmetry, e.g., rotations of multiples of 90 °
  - learned equivariance acts only on the training data and may not be as robust as equivariance by design
  - whether we can always expect to achieve system equivariance and/or whether it is always desirable （6 and 9)
  - There are also many unanswered theoretical questions in terms of both generalization and identifiability.

- **General group actions**
  - An interesting challenge is to account for group actions beyond rigid transformations, such as translations and rotations.

- **Beyond Euclidean domains**
  - The focus of the literature so far has been on scalar-valued imaging. Extensions to either the domain or range being a manifold or a graph are challenging and fall within the emerging framework of geometricdeep learning.