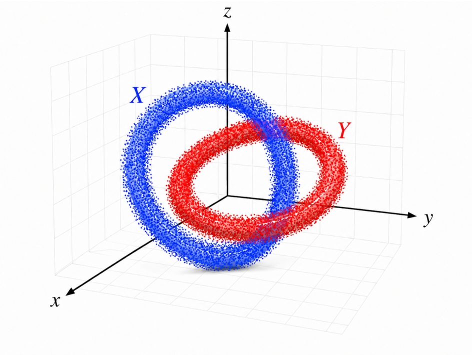
Low-dimensional topology of deep neural networks

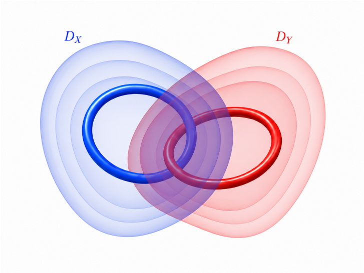
Junyu Ren • Lek-Heng Lim • University of Chicago • ICML 2026



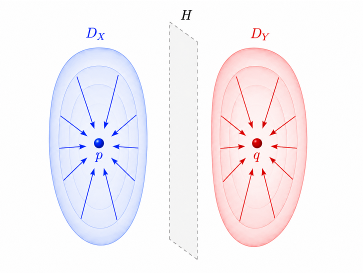
Can a network classify two **interlocked rings**?

A narrow **feedforward** ReLU net can **deform** them but not **unlink** them — it must **fold**.

Classification needs unlinking



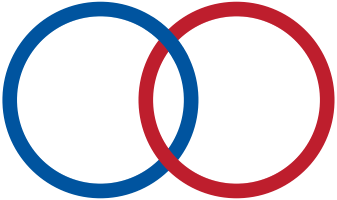
linked ⇒ **convex** decision regions intersect



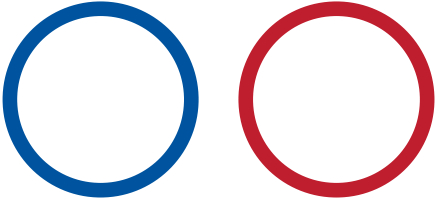
hyperplane-separated ⇒ **not** linked

Linear final layer ⇒ the net must **unlink** a class pair before it can separate it.

The obstruction: the linking number



Hopf link: $\mathrm{link} = \mathbf{1}$



Unlink: $\mathrm{link} = \mathbf{0}$

The linking number is an integer — as the classes deform continuously, it changes **only when they intersect**.

Can the net turn $\mathrm{link} = 1$ into 0?

$\Leftrightarrow$ can the transformation each layer expresses be **interpolated by a continuous deformation** of both classes **avoiding mutual collision**?

The impossibility theorem

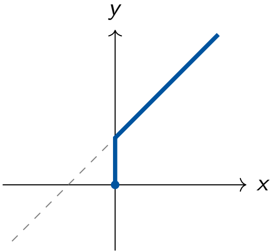
Monotone layers — invertible affines and squashing activations — **preserve the linking number** at every step.

Theorem. If two class manifolds are **linked** ($\mathrm{link} \neq 0$), then *no* width-d feedforward network with coordinate-wise **monotone** activations — ReLU, sigmoid, tanh, ... — can make them linearly separable, at **any depth**.

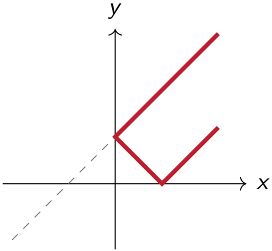
Perfect classification of linked classes is impossible. (Full proof in the paper.)

Folding — the geometric operation

Apply an activation coordinate-wise to the line $y = x + 1$:



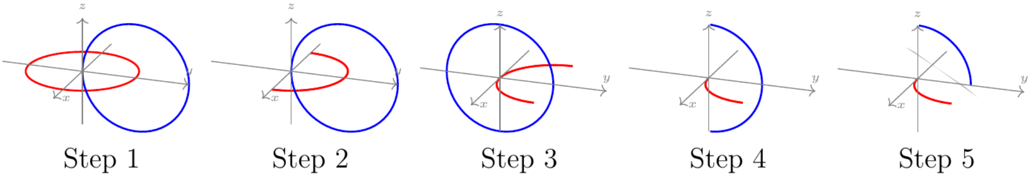
ReLU only **bends**: negatives clamp to 0



$|\cdot|$ **folds**: reflects $\pm x$ onto each other

A monotone activation can only **bend**; a **fold** reflects the two sides of an axis together — the move it can't make.

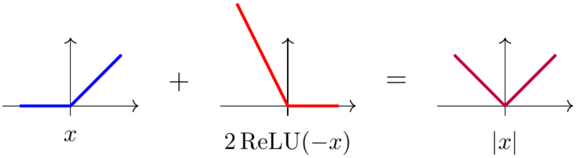
Folding — unlinking the Hopf link



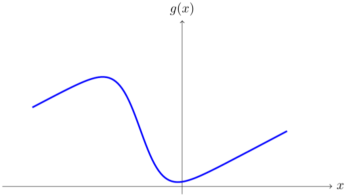
Interleave affine shifts with coordinate-wise $|\cdot|$ folds — the Hopf link **untangles**, until a plane separates the two curves.

Folding — the architectural escape

Which standard architectures **supply the fold**?



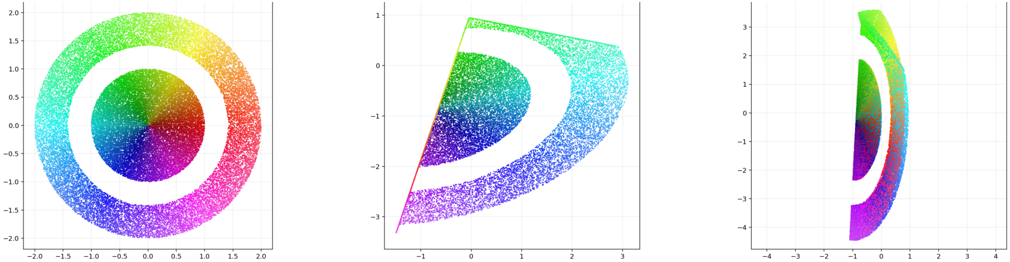
ResNet skip : $|x| = x + 2\,\mathrm{ReLU}(-x)$ in one residual block



Attention : a smoothed V-shape, locally like $|x|$

GeLU & Swish fold directly ; widening past the input dimension also works.

Experiments confirm the theory

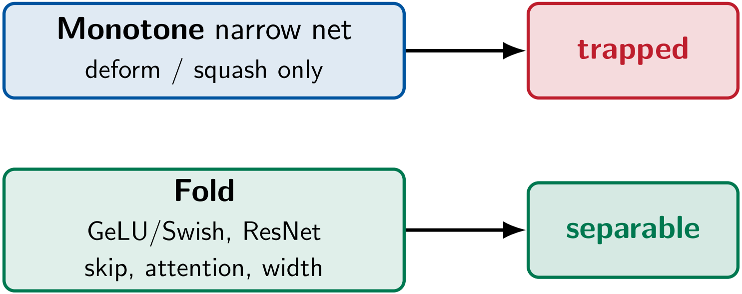


input x **residual fold** $\approx 2\,\mathrm{ReLU}(-x)$ **output** $x+f(x) = |x|$

A ResNet **visibly learns the fold**: the disk is peeled apart from the annulus.

ReLU stalls at 93% at *any* depth • **GeLU & ResNet reach 100%**

Topology: a lens on architecture



Topology asks which *shapes* a network can transform — not just which functions.

Poster: full theorem, attention & width results, and a suggestive real-data signal in CIFAR-10.