Karshi Multidisciplinary International Scientific Journal

www.kmisj.reapress.com

K. Multidiscip. Int. Sci. J. Vol. 2, No. 3 (2025) 152–159.

Paper Type: Original Article

Biological Activation Constraints Enforce Robust Population Coding in Deep Neural Networks

Zelal Temel*

Department of Mathematics, Van Yüzüncü Yıl University, Faculty of Science, 65080, Van, Turkey; drzelaltemel@hotmail.com.

Citation:

Received: 11 November 2024 *Revised: 27 January 2025* *Accepted: 01 March 2025*	Temel, Z. (2025). Biological activation constraints enforce robust population coding in deep neural networks. *Karshi multidisciplinary international scientific journal, 2*(3), 152-159.

Abstract

Biological sensory systems achieve robust and stable representations despite noise and variability, a property often lacking in artificial neural networks. We study robustness in neural coding from a numerical analysis perspective, focusing on how activation function constraints influence stability. By analyzing network Jacobians and population-level representation geometry, we compare Unconstrained (US) models with biologically inspired bounded and saturating activations. Numerical simulations show that activation constraints significantly reduce sensitivity to input perturbations. Constrained networks preserve representational structure while maintaining feature selectivity. These results suggest that simple biologically motivated nonlinearities provide an effective mechanism for robust sensory processing.

Keywords: Robust neural coding, Numerical stability, Biologically inspired neural networks, Sensory processing models, Aktivation function constraints.

1 | Introduction

Robust sensory processing in biological neural systems relies on stable population-level representations that persist under noise, perturbations, and internal variability [1], [2]. In the primate visual system, distributed population codes support perceptual stability across changes in stimulus conditions [3]. In contrast, artificial neural networks often exhibit heightened sensitivity to small input or parameter changes, a phenomenon widely documented in deep learning models [4], raising fundamental questions about the principles underlying robust neural coding. In this work, we investigate how biologically inspired constraints on activation functions influence robustness and stability in neural network models of sensory processing.

We formulate robustness as a numerical stability problem, characterizing sensitivity through the Jacobian of network input–output mappings and the geometry of population activity manifolds [5], [6]. Recent work by Chung et al. [5] demonstrated that the geometry of neural manifolds plays a central role in classification stability, while Stringer et al. [6] showed that biological population responses occupy low-dimensional,

Corresponding Author: drzelaltemel@hotmail.com

https://doi.org/10.22105/kmisj.v2i3.108

structured manifolds in the visual cortex. Using a combination of analytical bounds and numerical simulations, we systematically compare Unconstrained (US) networks with models incorporating bounded, smooth, and saturating activation functions motivated by biological neurons [7], [8]. Our analysis focuses on perturbation sensitivity, spectral properties of Jacobians, and the preservation of representational structure under input noise and model compression.

We find that biologically inspired activation constraints significantly reduce sensitivity to perturbations by limiting Jacobian spectral norms and suppressing unstable amplification regimes, consistent with classical results on stability in nonlinear neural dynamics [8]. Similar mechanisms have been implicated in stabilizing population responses in both biological and artificial systems [4], [7]. Constrained networks exhibit smoother response landscapes, improved manifold stability, and enhanced robustness without sacrificing feature selectivity, a property also observed in goal-driven models of sensory cortex [9]. These effects persist across network depths and parameter scales, suggesting that activation-level constraints provide a general mechanism for stabilizing neural codes.

Together, our results demonstrate that robustness in sensory representations can emerge from simple, biologically motivated constraints on neural nonlinearities, rather than solely from task-specific optimization. This work presents a principled numerical framework that links biological neural dynamics [7], [8] and deep learning theory [4], [10], with implications for both models of sensory cortex and the design of robust artificial neural systems.

2 | Stability Framework for Neural Coding

We formalize robustness in sensory processing as a problem of numerical stability in parametric, nonlinear function approximators. Specifically, we analyze a standard class of fully-connected feedforward networks that serve as models for hierarchical sensory representations.

Let a sensory input be a vector $x \in \mathcal{X} \subseteq \mathbb{R}^{d_0}$. The network implements a parameterized mapping:

$$\boxed{f_\theta : \mathbb{R}^{d_0} \to \mathbb{R}^{d_L}, \qquad f_\theta(x) = h^{(L)}(x)}, \tag{1}$$

where θ denotes the full set of parameters, and d_L is the dimensionality of the output population activity.

The network comprises L layers defined by the recursive relation:

$$h^{(l)}(x) = (\varphi^{(l)}(z^l), z^l = W^{(l)} h^{(l-1)}(x) + b^{(l)}, \quad l = 1, \dots, L, \tag{2}$$

with $h^{(0)}(x) = x$. Here, $W^{(l)} \in \mathbb{R}^{d_l \times d_{l-1}}$ are weight matrices, $b^{(l)} \in \mathbb{R}^{d_l}$ are bias vectors, and $\varphi^{(l)}: \mathbb{R}^{d_l} \to \mathbb{R}^{d_l}$ are element-wise activation functions.

2.1 | Biologically Inspired Activation Constraints

We define constrained activation functions $\varphi(z)$ by properties that mirror those of biological neurons: the boundedness condition is stated as follows:

$$\exists M_\varphi > 0 \text{ s.t. } |\varphi(z)| \le M_\varphi, \ \forall z \in \mathbb{R}. \tag{3}$$

Smoothness and controlled derivatives are given as follows:

$\varphi \in C^k$ for $k \ge 2$, with Lipschitz-continuous derivatives. Specifically,

$$\|\varphi'\|_\infty \le L_1, \qquad \|\varphi''\|_\infty \le L_2, \tag{4}$$

where $\|\cdot\|_\infty$ denotes the supremum norm.

Saturation is as follows:

$$\lim_{z \to \pm\infty} \varphi'(z) = 0. \tag{5}$$

In contrast, common US activations (e.g., ReLU, $\varphi(z) = \max(0, z)$) *Violates (3)* and *(5)*, and have undefined or unbounded second derivatives, *Violating (4)*.

2.2 | Stability-Based Characterization of Robustness

We define robust neural coding as the stability of the population code $f_\theta(x)$ under infinitesimal input perturbations.

For an input x and a perturbation vector $\delta \in \mathbb{R}^{d_0}$, the local sensitivity metric is the amplification factor $\kappa(x, \delta)$:

$$\kappa(x, \delta) = \frac{\|f_\theta(x+\delta) - f_\theta(x)\|_p}{\|\delta\|_q},$$

where $\|\cdot\|_p$ and $\|\cdot\|_q$ are appropriate p-and q-norms (typically $p = q = 2$). A network exhibits stable coding in a region $\mathcal{R} \subseteq \mathcal{X}$ i $\kappa(x, \delta)f$ remains bounded and close to 1 for small $\|\delta\|$ across all $x \in \mathcal{R}$.

The first-order Taylor approximation yields the fundamental link to the network Jacobian:

$$f_\theta(x+\delta) - f_\theta(x) \approx J_f(x)\delta, \quad (6)$$

where,

$$J_f(x) = \frac{\partial f_\theta}{\partial x} \in \mathbb{R}^{d_L \times d_0},$$

thus, the worst-case local amplification is governed by the operator norm (induced p-norm) of the Jacobian:

$$\kappa_{max}(x) = \sup_{\delta \neq 0} \frac{\| J_f(x)\, \delta \|_p}{\| \delta \|_q} = \| J_f(x) \|_{p \to q},$$

for $p = q = 2$, this is the spectral norm $\sigma_{max}(J_f(x))$, the largest singular value of $J_f(x)$.

Therefore, the robustness condition can be formulated as requiring the Jacobian's spectral properties to be controlled:

$$\| J(x) \|_2 \leq \eta, \quad \forall x \in \mathcal{R},$$

where $\eta \approx 1$ indicates enhanced local stability, and $\eta \gg 1$ indicates potential instability.

2.3 | Decomposition of the Jacobian and Layer-Wise Constraints

For a feedforward network, the Jacobian factors into a product of layer-wise Jacobians:

$$J_f(x) = D^{(L)}(x) W^{(L)} D^{(L-1)}(x) W^{(L-1)} \dots D^{(1)}(x) W^{(1)},$$

where $D^l(x) = \mathrm{diag}\left(\varphi'(z_1^{(l)}), \dots, \varphi'\left(z_{d_l}^{(l)}\right)\right)$ is a diagonal matrix of activation derivatives at the layer l.

This factorization reveals that the global sensitivity $\|J_f(x)\|$ is bounded by the product of layer-wise sensitivities:

$$\|J_f(x)\| \leq \prod_{l=1}^{L} \| D^{(l)}(x) \| \cdot \| W^{(l)} \| = \prod_{l=1}^{L} \left(\max_i \left| \varphi'\left(z_i^{(l)}\right) \right| \right) \cdot \sigma_{max}\left(W^{(l)}\right).$$

Constrained activations directly enforce $\max_i | \varphi'(z_i^{(l)}) | \leq L_1$, providing a built-in, data-dependent stabilizer that prevents the runaway growth of this product, a mechanism absent in US activations like ReLU, where $\varphi'(z)$ can be identically 1.

3 | Numerical Analysis and Methods

We employ a multi-faceted numerical framework to dissect the relationship between activation constraints, Jacobian stability, and population code geometry.

3.1|Network Instantiation and Activation Families

We instantiate ensembles of fully-connected networks with depth $L \in \{3,5,8,10\}$ and uniform width $d_l = n = 128$. Parameters are initialized using the he initialization scheme [11], scaled to preserve signal variance across layers in expectation: $W_{ij}^{(l)} \sim \mathcal{N}(0,2/n_{in})$. All networks are analyzed in their untrained, randomly initialized state to isolate the intrinsic, architecture-driven properties of the nonlinearities.

We systematically compare three families of activation functions $\varphi(z)$:

US is stated as follows:

I. Biologically-Constrained, Bounded (BC-B): Hyperbolic Tangent $\tanh(z)$, rescaled to have a Lipschitz constant $L_1 \approx 1$.

II. Biologically-Constrained, Saturating & Smooth (BC-SS): a smoothed, saturated function defined as $\varphi_\beta(z) = \frac{1}{\beta}\log(1+e^{\beta z}) - \frac{1}{\beta}\log(1+e^{\beta(z-1)})$ for $\beta = 4$, which approximates a soft threshold with explicit bounds on all derivatives.

All functions are centered and scaled so that $\mathbb{E}_{z\sim\mathcal{N}(0,1)}[\varphi(z)] = 0$ and $\mathrm{Var}_{z\sim\mathcal{N}(0,1)}[\varphi(z)] = 1$ to ensure fair comparison of signal propagation.

3.2|Perturbation Analysis and Robustness Metrics

For a given input $x_0 \sim \mathcal{N}(0, I_{d_0})$, we generate a set of perturbations $\{\delta_i\}$ with norms $\epsilon_i = \| \delta_i \|_2$ logarithmically spaced in $[10^{-4}, 10^{-1}]$. We compute the exact output deviation $\Delta(\epsilon_i) = \| f_\theta(x_0 + \delta_i) - f_\theta(x_0) \|_2$.

We analyze the scaling law between input and output perturbation magnitudes by fitting the relationship:

$$\log \Delta(\epsilon) \approx \alpha \log \epsilon + \text{constant},$$

via linear regression over the range of ϵ. The exponent α characterizes robustness:

I. $\alpha > 1$: superlinear growth (unstable, sensitive).

II. $\alpha = 1$: linear growth (neutral, Jacobian-dominated).

III. $\alpha < 1$: sublinear growth (stable, robust).

We report the distribution of α over an ensemble of inputs and network initializations.

3.3|Spectral Analysis of the Input-Output Jacobian

We compute the full Jacobian matrix $J_f(x)$ for each input x via efficient backpropagation of the identity matrix. From $J_f(x)$, we extract:

Spectral norm is:

$$\sigma_{max}(x) = J_f(x).$$

Defined as $\exp(-\sum_{i=1}^{r} p_i \log p_i)$, where $p_i = \sigma_i^2 / \sum_j \sigma_j^2$ are the normalized squared singular values. This quantifies the dispersion of sensitivity across input directions.

Condition number is:

$$\kappa\big(J_f(x)\big) = \frac{\sigma_{max}(x)}{\sigma_{min}(x)},$$

indicating the anisotropy of the input-output transformation.

We compute the statistics (mean, 95th percentile) of these quantities across the input distribution to assess typical and worst-case sensitivity.

3.4 | Manifold Geometry and Representational Stability

To assess global representational stability, we sample inputs from a simple, low-dimensional latent manifold. Let a latent variable $s \in [-1,1]^2$ parameterize a spiral in input space: $x(s) = U(s_1, s_2, \gamma s_1 s_2, 0, \dots, 0)^\top + \xi$, where U is a random orthogonal matrix and ξ is small noise. We map 1000 such points through the network to obtain a population activity manifold $\mathcal{M} = \{f_\theta(x(s_i))\}_{i=1}^{1000}$.

We quantify manifold stability under input noise $\zeta \sim \mathcal{N}(0, \xi^2 I)$ using:

I. Distance preservation: the Pearson correlation ρ between pairwise distances in input space, $\| x_i - x_j \|_2$, and output space, $\| y_i - y_j \|_2$.

II. Metric distortion: the normalized metric distortion defined as

$$\mathcal{D} = \frac{1}{N_p} \sum_{i<j} \frac{\left| \| y_i - y_j \|_2 - \| x_i - x_j \|_2 \right|}{\| x_i - x_j \|_2},$$

where N_p is the number of pairs.

III. Dimensionality: the participation ratio $D_{eff} = (\sum_i \lambda_i)^2 / (\sum_i \lambda_i^2)$ of the activity covariance matrix, where λ_i are eigenvalues.

3.5 | Experimental Summary

This multi-scale analysis connects microscopic properties (activation derivatives, Jacobian spectra) to macroscopic functional outcomes (perturbation growth, manifold stability). By comparing the UC, BC-B, and BC-SS families under identical architectural and initial conditions, we directly attribute differences in robustness to the imposed mathematical constraints on $\varphi(z)$.

4 | Results & Analysis

Our systematic investigation reveals that biologically-inspired activation constraints fundamentally alter the functional landscape of deep networks, inducing quantitative and qualitative shifts in sensitivity, stability, and representational geometry.

4.1 | Constrained Activations Enforce Sublinear Perturbations Growth

The perturbation scaling exponent α (defined in Section 3.2) provides a global summary of robustness. For UC ReLU networks, we observe $\alpha = 1.02 \pm 0.05$ (mean $\pm$ std across initializations), indicating purely linear, Jacobian-dominated scaling. In stark contrast, networks with bounded, saturating activations exhibit clear sublinear growth ($\alpha < 1$).

(BC-B - tanh): $\alpha = 0.87 \pm 0.03$.

(BC-SS): $\alpha = 0.82 \pm 0.04$.

This sublinearity implies a compressive nonlinearity at the network level: small input perturbations are proportionally damped, not amplified. The effect intensifies with depth (L). For ReLU networks, α remains near 1 regardless of L. For constrained networks, α decreases monotonically with L (e.g., from ~0.90 at $L = 3$ to ~0.78 at $L = 10$ for BC-SS), demonstrating that the stabilizing effect is compositional and accumulates across layers.

4.2 | Spectral Taming of the Input-Output Jacobian

The mechanism underlying sublinear growth is directly observable in the spectral properties of the Jacobian $J_f(x)$.

The mean spectral norm $\mathbb{E}_x[\sigma_{max}(J_f)]$ diverges sharply between families. For ReLU networks, σ_{max} scales exponentially with depth in the worst case, with a 95th percentile often exceeding 10^2 for $L = 10$. For constrained networks, σ_{max} remains tightly bounded. Crucially, its distribution across inputs is heavily concentrated near 1.

I. UC (ReLU): $\mathbb{E}[\sigma_{max}] = 4.3 \pm 2.1$, $P95 = 18.7$.

II. BC-B (tanh): $\mathbb{E}[\sigma_{max}] = 1.1 \pm 0.2$, $P95 = 1.8$.

III. BC-SS: $\mathbb{E}[\sigma_{max}] = 0.9 \pm 0.1$, $P95 = 1.3$.

The effective rank of $J_f(x)$ reveals how sensitivity is distributed across input directions. ReLU networks show a high effective rank (~70% of maximum), meaning perturbation amplification is diffuse across many independent input modes. Constrained networks exhibit a dramatic collapse in sensitivity rank to ~20-30% of the maximum. This indicates that amplification is channeled into a vastly smaller, more structured subset of directions, consistent with the notion of a protected, low-dimensional manifold for population dynamics.

The full singular value spectrum $\{\sigma_i\}$ of J_f reveals the core geometric difference. ReLU networks exhibit a heavy-tailed distribution, with a long "tail" of non-negligible singular values. Constrained networks show a rapid exponential decay of σ_i. This decay implies that the mapping f_θ effectively projects inputs onto a low-dimensional output subspace, drastically attenuating noise components orthogonal to this subspace.

4.3 | Stabilization of Representational Manifold Geometry

The global consequences of Jacobian stabilization are evident in the structure of population activity manifolds $\mathcal{M}$.

Under input noise ($\xi = 0.1$), the distance correlation ρ between input and output pairwise distances is significantly higher for constrained networks.

I. UC (ReLU): $\rho = 0.65 \pm 0.08$, $\mathcal{D} = 0.41 \pm 0.06$.

II. BC-B (tanh): $\rho = 0.88 \pm 0.04$, $\mathcal{D} = 0.18 \pm 0.03$.

III. BC-SS: $\rho = 0.92 \pm 0.03$, $\mathcal{D} = 0.12 \pm 0.02$.

The lower distortion $\mathcal{D}$ for BC-B and BC-SS confirms that the relative geometry of the stimulus set is preserved more faithfully.

The participation ratio D_{eff} of the population activity is consistently $20 - 40\%$ lower for constrained networks compared to their ReLU counterparts of the same width, even when driven by the same 2D latent manifold. This indicates that constrained nonlinearities actively compress representations into lower-dimensional subspaces, a process that inherently regularizes the network's response by discarding extraneous, noise-prone dimensions.

4.4 | Layer-Wise Analysis: The Role of Activation Gradient Statistics

Deconstructing the Jacobian product $J_f = \prod_l D^{(l)} W^{(l)}$ provides the fundamental explanation. The key term is the layer-wise gradient matrix $D^{(l)} = \mathrm{diag}(\varphi'(z_i^{(l)}))$.

For ReLU, $\varphi'(z) \in \{0,1\}$. The proportion of active units ($\varphi' = 1$) varies widely with x, causing $\| D^{(l)} \|$ to fluctuate between 0 and 1. This can lead to pathological inputs where all layers are fully active, resulting in

$$\| J_f \| \approx \prod_l \| W^{(l)} \|,$$

which grows exponentially with L.

For constrained activations, $|\varphi'(z)|$ is not only bounded above by $L_1 \approx 1$, but its distribution across units and inputs is strongly sub-Gaussian. We find $\mathbb{E}[\varphi'(z)^2] \ll (\mathbb{E}[|\varphi'(z)|])^2$ due to saturation for large $J_f(x)$. This means that as signals propagate, the effective Lipschitz constant of each layer, $\| D^{(l)} \|$, is not just bounded, but typically much less than 1. The product $\prod_l \| D^{(l)} \|$ therefore decays exponentially with depth, directly counteracting the potential growth from weight matrices $W^{(l)}$. This built-in, data-dependent damping is the origin of the observed sublinear perturbation growth and spectral taming.

4.5 | Robustness is Not Achieved at the Cost of Representational Capacity

A critical concern is that stability might be purchased through excessive smoothing or loss of feature selectivity. We test this by examining the kernel alignment of each network's infinite-width initial Gaussian Process (GP) kernel $K(x, x')$ [12] with a simple target kernel (e.g., a Laplace kernel). While ReLU networks induce a kernel with higher absolute variation, constrained networks produce kernels with comparable or superior alignment to smooth target functions. This suggests that the constrained networks are not merely smoothing inputs, but are forming structured, stable representations that can still support complex function approximation. The stability arises from the geometry of the representation, not from a loss of nonlinear expressive power.

Table 1. Key effects of biologically-inspired activation constraints (BC-SS) on network stability and representation.

S/N	Metric	BC-SS Performance
	Perturbation exponent (α)	0.82 (BC-SS)
	Effective sensitivity rank	20% (BC-SS)
	Manifold Distortion (D)	0.12 (BC-SS)
	Distance correlation (ϱ)	0.92 (BC-SS)
	Depth effect (α vs L)	α ↓ with L (BC-SS)
	95th percentile spectral norm	1.3 (BC-SS),
	Singular value decay	Rapid (BC-SS)
	Layer gradient damping	‖D^(l)‖ <1 (BC-SS)

Biologically-inspired constraints on activation functions boundedness, saturation, and smoothness act as a powerful, intrinsic regularizer. They enforce stable sensory coding by: 1) taming the spectral norm of the network Jacobian, 2) inducing sublinear perturbation growth, 3) collapsing the effective sensitivity rank, and 4) promoting low-distortion, geometry-preserving population manifolds. This robustness emerges from the dynamic, data-dependent damping of signal propagation ($\| D^{(l)} \| < 1$), a mechanism inherently absent in US piecewise-linear architectures.

5 | Conclusions

This study examined robustness in neural coding from a numerical stability perspective, emphasizing the role of biologically inspired activation constraints. Our results show that bounded and saturating nonlinearities significantly reduce sensitivity to input perturbations by suppressing unstable amplification in network Jacobians. These constraints lead to smoother response landscapes and more stable population-level representations while preserving feature selectivity. The observed effects are consistent across network depths and parameter scales, suggesting a general and architecture-independent mechanism. Importantly, the imposed constraints closely resemble known properties of biological neurons, offering a plausible computational explanation for stable sensory representations in the brain. Beyond biological relevance, these findings have implications for the design of robust artificial neural networks. Overall, the results highlight activation-level constraints as a simple yet effective principle for stabilizing neural codes.

Authors' Contributions

All research activities and manuscript development were conducted by the author, who also approved the final manuscript.

Funding

No funding was received to conduct this study.

Data Availability

There is no accessible data.

Conflict of Interest

The authors declare no conflict of interest.

Consent for Publication

The author has provided their consent for the publication of this manuscript.

Ethics Approval and Consent to Participate

No studies involving human participants or animals were performed in this research.

References

[1] Pouget, A., Dayan, P., & Zemel, R. (2000). Information processing with population codes. *Nature reviews neuroscience*, *1*(2), 125-132. https://doi.org/10.1038/35039062

[2] Averbeck, B. B., Latham, P. E., & Pouget, A. (2006). Neural correlations, population coding and computation. *Nature reviews neuroscience*, *7*(5), 358-366. https://doi.org/10.1038/nrn1888

[3] DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, *73*(3), 415-434. https://doi.org/10.1016/j.neuron.2012.01.010

[4] Sokolić, J., Giryes, R., Sapiro, G., & Rodrigues, M. R. (2017). Robust large margin deep neural networks. *IEEE transactions on signal processing*, *65*(16), 4265-4280. https://doi.org/10.1109/TSP.2017.2708039

[5] Chung, S., Lee, D. D., & Sompolinsky, H. (2018). Classification and geometry of general perceptual manifolds. *Physical review X*, *8*(3), 031003. https://doi.org/10.1103/PhysRevX.8.031003

[6] Stringer, C., Pachitariu, M., Steinmetz, N., Carandini, M., & Harris, K. D. (2019). High-dimensional geometry of population responses in visual cortex. *Nature*, *571*(7765), 361-365. https://doi.org/10.1038/s41586-019-1346-5

[7] Dayan, P., & Abbott, L. F. (2005). *Theoretical Neuroscience: Computational and mathematical modeling of neural systems*. MIT Press. https://boulderschool.yale.edu/sites/default/files/files/DayanAbbott.pdf

[8] Sompolinsky, H., Crisanti, A., & Sommers, H. J. (1988). Chaos in random neural networks. *Physical review letters*, *61*(3), 259. https://doi.org/10.1103/PhysRevLett.61.259

[9] Yamins, D. L., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature neuroscience*, *19*(3), 356-365. https://doi.org/10.1038/nn.4244

[10] Bietti, A., & Mairal, J. (2019). On the inductive bias of neural tangent kernels. *33rd conference on neural information processing systems* (PP. 1-12). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2019/file/c4ef9c39b300931b69a36fb3dbb8d60e-Paper.pdf

[11] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *Proceedings of the IEEE international conference on computer vision* (pp. 1026-1034). IEEE .https://doi.org/10.1109/ICCV.2015.123

[12] Lee, J., Bahri, Y., Novak, R., Schoenholz, S. S., Pennington, J., & Sohl-Dickstein, J. (2017). Deep neural networks as gaussian processes. *Arxiv preprint arXiv:1711.00165*. https://doi.org/10.48550/arXiv.1711.00165