Supplemental Material

A. Discussion of Mean Field Theory

Consider an L-layer 1D[4] periodic CNN with filter size $2k+1$, channel size c, spatial size n, per-layer weight tensors $\omega \in \mathbb{R}^{(2k+1)\times c\times c}$ and biases $b \in \mathbb{R}^c$. Let $\phi : \mathbb{R} \to \mathbb{R}$ be the activation function and let $h_j^l(\alpha)$ denote the pre-activation at layer l, channel j, and spatial location α. Suppose the weights ω_{ij}^l are drawn i.i.d. from the Gaussian $\mathcal{N}(0, \sigma_\omega^2/(c(2k+1)))$ and the biases b_j^l are drawn i.i.d. from the Gaussian $\mathcal{N}(0, \sigma_b^2)$. The forward-propagation dynamics can be described by the recurrence relation,

$$h_j^{l+1}(\alpha) = \sum_{i\in chn}\sum_{\beta\in ker} x_i^l(\alpha+\beta)\omega_{ij}^{l+1}(\beta) + b_j^{l+1}\,, \quad x_i^l(\alpha) = \phi(h_i^l(\alpha)).$$

For $l \geq 0$, note that (a) $\{h_j^{l+1}\}_j$ are i.i.d. random variables and (b) for each j, h_j^{l+1} is a sum of c i.i.d. random variables with mean zero. The central limit theorem implies that $\{h_j^{l+1}\}_j$ are i.i.d. Gaussian random variables. Let $\mathbf{\Sigma}^{l+1} = \{\Sigma^{l+1}_{\alpha,\alpha'}\}_{\alpha,\alpha'}$ denote the covariance matrix, where

$$\Sigma^{l+1}_{\alpha,\alpha'} = \mathbb{E}\left[h_j^{l+1}(\alpha)h_j^{l+1}(\alpha')\right],$$

where the expectation is taken over all random variables in and before layer $(l+1)$. Therefore, we have the following lemma.

Lemma A.1. *As $c \to \infty$, for each $l \geq 0$, h_j^{l+1} is a mean zero Gaussian with covariance matrix $\mathbf{\Sigma}^{l+1}$ satisfying the recurrence relation,*

$$\mathbf{\Sigma}^{l+1} = \mathcal{A} \star \mathcal{C}(\mathbf{\Sigma}^l)\,. \tag{S1}$$

Proof. Let $\theta^l = [W^l, b^l]$ and $\theta^{0:l} = [\theta^0, \ldots, \theta^l]$. Then,

$$\Sigma^{l+1}_{\alpha,\alpha'} = \mathbb{E}_{\theta^{0:l}}\,\mathbb{E}_{\theta^{l+1}}\left[h_j^{l+1}(\alpha)h_j^{l+1}(\alpha')\right] \tag{S2}$$

$$= \mathbb{E}_{\theta^{0:l}}\left[\frac{\sigma_\omega^2}{2k+1}\sum_{\beta\in ker}\frac{1}{c}\sum_{i=1}^{c}(x_i^l(\alpha+\beta)x_i^l(\alpha'+\beta)) + \sigma_b^2\right], \tag{S3}$$

where we used the fact that,

$$\mathbb{E}_{\theta^{l+1}}\left[\omega_{ij}^{l+1}(\beta)\omega_{i'j'}^{l+1}(\beta')\right] = \begin{cases} \frac{\sigma_\omega^2}{c(2k+1)}, & \text{if} \quad (i,j,\beta) = (i',j',\beta') \\ 0, & \text{otherwise.} \end{cases}$$

Note that $\mathbf{\Sigma}^1$ can be computed once h^0 (or x^0) is given. We will proceed by induction. Let $l \geq 1$ be fixed and assume $\{h_j^l\}_j$ are i.i.d. mean zero Gaussian with covariance $\mathbf{\Sigma}^l$. It is not difficult to see that $\{h_j^{l+1}\}_j$ are also i.i.d. mean zero Gaussian as $c \to \infty$. To compute the covariance, note that for any fixed pair (α, α'), $\{x_i(\alpha)x_i(\alpha')\}_i$ are i.i.d. random variables. Then,

$$\mathbb{E}_{\theta^{0:l}}\left[\frac{1}{c}\sum_{i=1}^{c}(x_i^l(\alpha)x_i^l(\alpha'))\right] = \mathbb{E}_{\theta^{0:l}}\left[x_i^l(\alpha)x_i^l(\alpha')\right] = \mathbb{E}_{\theta^{0:l}}\left[\phi(h_i^l(\alpha))\phi(h_i^l(\alpha'))\right]. \tag{S4}$$

Thus by eq. (2.5), eq. (S2) can be written as,

$$\Sigma^{l+1}_{\alpha,\alpha'} = \frac{1}{2k+1}\sum_{\beta\in ker}[C(\mathbf{\Sigma}^l)]_{\alpha+\beta,\alpha'+\beta}\,, \tag{S5}$$

so that,

$$\mathbf{\Sigma}^{l+1} \equiv \mathcal{A} \star \mathcal{C}(\mathbf{\Sigma}^l)\,. \tag{S6}$$

□

[4]For notational simplicity, as in the main text, we again consider 1D convolutions, but the 2D case proceeds identically.

The same proof yields the following corollary.

Corollary A.2. *Let $v = (v_\beta)_{\beta \in ker}$ be a sequence of non-negative numbers with $\sum_{\beta \in ker} v_\beta = 1$. Let $\mathcal{A}_v$ be the cross-correlation operator induced by v, i.e.,*

$$(\mathcal{A}_v \star f)_{\alpha,\alpha'} = \sum_{\beta \in ker} v_\beta f_{\alpha+\beta,\alpha'+\beta}\,. \tag{S7}$$

Suppose the weights $\omega^l_{i,j}(\beta)$ are drawn i.i.d. from the Gaussian $\mathcal{N}(0, \frac{v_\beta}{c} \cdot \sigma^2_\omega)$. Then the recurrence relation for the covariance matrix is given by,

$$\mathbf{\Sigma}^{l+1} = \mathcal{A}_v \star \mathcal{C}(\mathbf{\Sigma}^l)\,. \tag{S8}$$

A.1. Back-propagation

Let E denote the loss associated to a CNN and $\delta^l_j(\alpha)$ denote a backprop signal given by,

$$\delta^l_j(\alpha) = \frac{\partial E}{\partial h^l_j(\alpha)}.$$

The layer-to-layer recurrence relation is given by,

$$\begin{aligned}\delta^l_j(\alpha) &= \sum_{i \in chn} \sum_{\alpha' \in sp} \frac{\partial E}{\partial h^{l+1}_i(\alpha')} \frac{\partial h^{l+1}_i(\alpha')}{\partial h^l_j(\alpha)} \\ &= \sum_{i \in chn} \sum_{\beta \in ker} \delta^{l+1}_i(\alpha-\beta)\omega^{l+1}_{ji}(\beta)\phi'(h^l_j(\alpha))\,.\end{aligned}$$

We need to make an assumption that the weights used during back-propagation are drawn independently from the weights used in forward propagation. This implies $\{\delta^l_j\}_{j \in chn}$ are independent for all l and for $j \neq j'$,

$$\mathbb{E}\left[\delta^l_j(\alpha)\delta^l_{j'}(\alpha')\right] = 0\,,$$

and

$$\begin{aligned}\mathbf{\Sigma}_{\alpha,\alpha'} &= \mathbb{E}\left[\delta^l_j(\alpha)\delta^l_j(\alpha')\right] \\ &= \sum_{i \in chn} \sum_{\beta \in ker} \mathbb{E}\left[\delta^{l+1}_i(\alpha-\beta)\delta^{l+1}_i(\alpha'-\beta)\right]\mathbb{E}\left[\phi'(h^l_j(\alpha)\phi'(h^l_j(\alpha')\right]\mathbb{E}\left[\omega^{l+1}_{j,i}(\beta)\omega^{l+1}_{j,i}(\beta)\right] \\ &= \left(\frac{1}{2k+1}\sum_{\beta \in ker} \mathbf{\Sigma}_{\alpha-\beta,\alpha'-\beta}\right)\left(\sigma^2_\omega \mathbb{E}\left[\phi'(h^l_j(\alpha)\phi'(h^l_j(\alpha'))\right]\right)\,.\end{aligned}$$

For large l, the second parenthesized term can be approximated by χ_1 if $\alpha' = \alpha$ and by χ_{c^*} otherwise.

B. The Jacobian of the $\mathcal{C}$-map

Recall that $\mathcal{C} : \mathrm{PSD}_n \to \mathrm{PSD}_n$ is given by,

$$\left[\mathcal{C}(\mathbf{\Sigma})\right]_{\alpha,\alpha'} = \sigma^2_\omega\, \mathbb{E}_{\boldsymbol{h} \sim \mathcal{N}(0,\mathbf{\Sigma})}\left[\phi(h_\alpha)\phi(h_{\alpha'})\right] + \sigma^2_b\,. \tag{S9}$$

We are interested in the linearized dynamics of $\mathcal{C}$ near the fixed point $\mathbf{\Sigma}^*$. Let $J : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ denote the Jacobian of $\mathcal{C}$ at $\mathbf{\Sigma}^*$. The main result of this section is that J commutes with any diagonal convolution operator.

Theorem B.1. *Let J be as above and $\mathcal{A}$ be any $n \times n$ diagonal matrix and U be any $n \times n$ symmetric matrix. Then,*

$$\mathcal{A} \star J(U) = J(\mathcal{A} \star U)\,. \tag{S10}$$

Let $\{V_{\alpha,\alpha'}\}_{0 \le \alpha \le \alpha' \le n-1}$ be the canonical basis of the space of $n \times n$ symmetric matrices, i.e. $[V_{\alpha,\alpha'}]_{\bar\alpha,\bar\alpha'} = 1$ if $(\alpha,\alpha') = (\bar\alpha,\bar\alpha')$ or $(\bar\alpha',\bar\alpha)$ and 0 otherwise. We claim the following:

Lemma B.2. *The Jacobian J has the following representation:*

- *For the off-diagonal terms (i.e. $\alpha \neq \alpha'$),*
$$JV_{\alpha,\alpha'} = \chi_{c^*} V_{\alpha,\alpha'}. \tag{S11}$$
- *For the diagonal terms,*
$$JV_{\alpha,\alpha} = \chi_{q^*} V_{\alpha,\alpha} + \kappa \sum_{\alpha' \neq \alpha} V_{\alpha,\alpha'}\,, \tag{S12}$$
where κ is given by,
$$\kappa = \frac{\sigma_\omega^2}{2} \mathbb{E}_{\boldsymbol{h}\sim\mathcal{N}(0,\boldsymbol{C}^*)} \phi(h_1)\phi''(h_2),\ h_1 \neq h_2\,. \tag{S13}$$

We first prove Theorem B.1 assuming Lemma B.2, and afterwards we prove the latter.

Proof of Theorem B.1. It is clear that
$$V_{\text{o.d.}} = \text{span}\{V_{\alpha,\alpha'} : \alpha \neq \alpha'\}$$
is an eigenspace of J with eigenvalue χ_{c^*}. Here $\text{span}\{X\}$ denotes the linear span of X. For $\chi_{q^*} \neq \chi_{c^*}$, define,
$$\tilde{V}_\alpha = V_{\alpha,\alpha} + \frac{\kappa}{\chi_{q^*} - \chi_{c^*}} \sum_{\alpha' \neq \alpha} V_{\alpha,\alpha'}.$$
It is straightforward to verify that
$$V_{\text{d}} = \text{span}\{\tilde{V}_\alpha : \alpha \in sp\}$$
is an eigenspace of J with eigenvalue χ_{q^*} and the direct sum $V_{\text{d}} \bigoplus V_{\text{o.d.}}$ is the whole space of $n \times n$ symmetric matrices. Note that J acts on $V_{\text{o.d.}}$ in a pointwise fashion and that $\mathcal{A}$ maps $V_{\text{o.d.}}$ onto itself (one can form an eigen-decomposition of $\mathcal{A}$ (and J) in $V_{\text{o.d.}}$ using Fourier matrices; see below for details.) Thus $\mathcal{A}$ commutes with J in $V_{\text{o.d.}}$. It remains to verify that they also commute in V_{d}.

,A key observation is that $\{\tilde{V}_\alpha\}_{\alpha\in sp}$ has a nice group structure,
$$\{\tilde{V}_\alpha\}_{\alpha\in sp} = \{V_{\alpha,\alpha} \star \tilde{V}_0\}_{\alpha\in sp}\,,$$
which we can use it to form a new basis for V_{d},
$$V_{\text{d}} = \text{span}\{U_\alpha : U_\alpha = F_\alpha \star \tilde{V}_0, \alpha \in sp\}\,, \tag{S14}$$
where F_α is the diagonal matrix formed by the α-th row of the $n \times n$ Fourier matrix, i.e. $F_\alpha = \text{diag}((f_{\alpha,\alpha'})_{\alpha'\in sp})$ with $f_{\alpha,\alpha'} = \frac{1}{\sqrt{n}} e^{2\pi i \alpha\alpha'\pi/n}$. Since each F_α is an eigen-vector of the $2D$ convolutional operator $\mathcal{A}\star\cdot$ ($\mathcal{A}$ is diagonal),
$$\mathcal{A}\star(JU_\alpha) = \chi_{q^*}\mathcal{A}\star U_\alpha = \chi_{q^*}\mathcal{A}\star F_\alpha\star\tilde{V}_0 = \chi_{q^*}(\mathcal{A}\star F_\alpha)\star\tilde{V}_0 = \chi_{q^*}\lambda_\alpha F_\alpha\star\tilde{V}_0 = \chi_{q^*}\lambda_\alpha U_\alpha = J(\mathcal{A}\star U_\alpha)$$
where λ_α is the eigenvalue of F_α. This finishes our proof. □

Proof of Lemma B.2. We first consider perturbing the off-diagonal terms. Let ϵ be a small number and $\alpha \neq \alpha'$. Note that for $(\bar{\alpha},\bar{\alpha}') \notin \{(\alpha,\alpha'),(\alpha',\alpha)\}$,
$$[\mathcal{C}(\boldsymbol{\Sigma}^* + \epsilon V_{\alpha,\alpha'})]_{\bar{\alpha},\bar{\alpha}'} = [\mathcal{C}(\boldsymbol{\Sigma}^*)]_{\bar{\alpha},\bar{\alpha}'} \tag{S15}$$
and
$$[\mathcal{C}(\boldsymbol{\Sigma}^* + \epsilon V_{\alpha,\alpha'})]_{\alpha,\alpha'} = \sigma_\omega^2 \mathbb{E}\phi(h_1)\phi(h_2) + \sigma_b^2, \tag{S16}$$
where $(h_1,h_2) \sim \mathcal{N}(0,Q)$ with $Q_{11} = Q_{22} = q^*$ and $Q_{12} = Q_{21} = c^*q^* + \epsilon$. Let $c = c^* + \epsilon/q^*$ and choose two independent random variables $u_1, u_2 \sim \mathcal{N}(0,1)$. Then,
$$[\mathcal{C}(\boldsymbol{\Sigma}^* + \epsilon V_{\alpha,\alpha'})]_{\alpha,\alpha'} = \sigma_\omega^2 \mathbb{E}\phi(\sqrt{q^*}u_1)\phi(\sqrt{q^*}(cu_1 + \sqrt{1-c^2}u_2)) + \sigma_b^2. \tag{S17}$$

Taylor expanding the term $\phi(\sqrt{q^*}(cu_1 + \sqrt{1-c^2}u_2))$ about the point $\sqrt{q^*}(c^*u_1 + \sqrt{1-(c^*)^2}u_2)$, one can show,

$$[\mathcal{C}(\mathbf{\Sigma}^* + \epsilon V_{\alpha,\alpha'})]_{\alpha,\alpha'} = c^*q^* + \chi_{c^*}\epsilon + O(|\epsilon|^2)\,, \tag{S18}$$

which proves the first statement of Lemma B.2.

To prove the second statement, let α be fixed and perturb $\mathbf{\Sigma}^*$ by $\epsilon V_{\alpha,\alpha}$. Note that all the terms are unchanged except the ones in the α-th row or α-th column. It is straightforward to show (see (Poole et al., 2016)) that

$$[\mathcal{C}(\mathbf{\Sigma}^* + \epsilon V_{\alpha,\alpha})]_{\alpha,\alpha} = q^* + \chi_{q^*}\epsilon + O(|\epsilon|^2). \tag{S19}$$

For any $\alpha' \neq \alpha$,

$$[\mathcal{C}(\mathbf{\Sigma}^* + \epsilon V_{\alpha,\alpha})]_{\alpha',\alpha} = \sigma_\omega^2 \mathbb{E}\phi(\sqrt{q^*}u_1)\phi(\sqrt{q^*}c^*u_1 + q(\epsilon)u_2)) + \sigma_b^2\,, \tag{S20}$$

where u_1 and u_2 are the same as in eq.(S17) and,

$$q(\epsilon) = \sqrt{q^* + \epsilon - (c^*)^2 q^*}\,. \tag{S21}$$

We can then Taylor expand $q(\epsilon)$ about $\sqrt{q^* - (c^*)^2q^*}$, $\phi(\sqrt{q^*}c^*u_1 + q(\epsilon)u_2))$ about $(\sqrt{q^*}c^*u_1 + \sqrt{q^* - (c^*)^2q^*}u_2)$, and apply one integration by parts to the second variable (namely, apply the identity $\mathbb{E}u_1 f(u_1) = \mathbb{E}f'(u_1)$) to find,

$$[\mathcal{C}(\mathbf{\Sigma}^* + \epsilon V_{\alpha,\alpha})]_{\alpha',\alpha} = c^*q^* + \kappa\epsilon + O(|\epsilon|^2). \tag{S22}$$

□

C. Construction of Random Orthogonal Kernels

C.1. Computational Complexity

For simplicity, consider constructing a $\Bbbk \times \Bbbk \times c \times c$ orthogonal kernel. The complexity can be roughly determined as follows:

1. Constructing $O(\Bbbk)$ $c \times c$ symmetric orthogonal matrices takes $O(\Bbbk c^3)$ steps.
2. For $j = 1, \ldots, \Bbbk - 1$, convolving a $j \times j$ (each entry is a $c \times c$ matrix) matrix with a 2×2 matrix requires $O(j^2)$ matrix multiplications between two $c \times c$ matrices. Since each matrix multiplication costs $O(c^3)$, a total number of $O((\Bbbk c)^3)$ steps is required for block-wise matrix convolutions.
3. In sum, the computational complexity is about $O((\Bbbk c)^3)$.

C.2. Delta Orthogonal Kernels

Algorithm 2 2D Delta orthogonal kernels for CNNs, available in TensorFlow via the ConvolutionDeltaOrthogonal initializer.

Input: $\Bbbk$ kernal size, c_{in} number of input channels, c_{out} number of output channels.
Return: a $\Bbbk \times \Bbbk \times c_{in} \times c_{out}$ tensor K
Step 1. Randomly generate a $c_{in} \times c_{out}$ matrix H with orthonormal rows.
Step 2. Define a $\Bbbk \times \Bbbk \times c_{in} \times c_{out}$ tensor K in the following way: for β, β' in $0, 1, \ldots, \Bbbk - 1$, for $i = 0, \ldots, c_{in} - 1$, $j = 0, \ldots, c_{out} - 1$, set

$$K(\beta, \beta', i, j) = \begin{cases} H(i,j), & \text{if } \beta = \beta' = [\Bbbk/2] \\ 0, & \text{otherwise.} \end{cases}$$

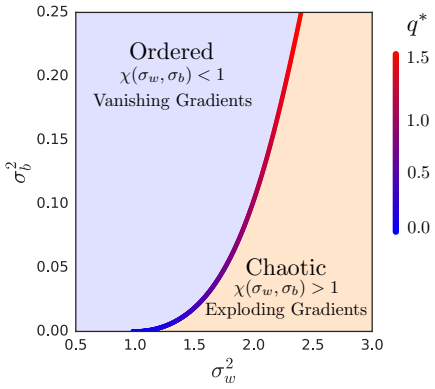

Figure S1. Phase diagram for fully-connected networks (reproduced from (Pennington et al., 2017)). In our analysis, we find they also apply to CNNs.

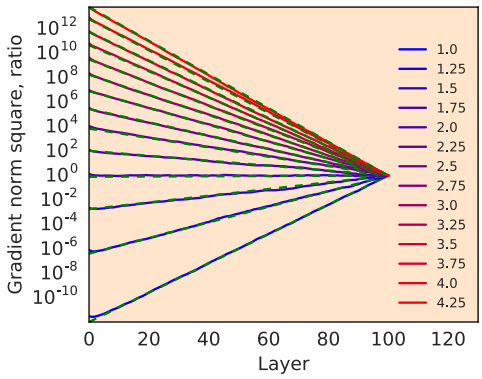

Figure S2. Vanishing/exploding gradients in a random CNN as the weight variance σ_w^2 is swept so as to transition from the ordered phase (blue curves) to the chaotic phase (red curves). Mean field theory predicts an exponential decay/growth, which is overlaid (green, dashed).

D. Phase Diagram and Vanishing/Exploding gradients

D.1. Phase Diagram

Figure S1 shows the phase diagram derived from the mean field theory of signal propagation in fully-connected networks, reproduced from (Pennington et al., 2017). It depicts the ordered and chaotic phases (with vanishing and exploding gradients, respectively) separated by a transition. The variation in value of q^* along the critical line is shown in color. As discussed in the main text, it also applies to the ordered-to-chaotic phase transition of CNNs.

D.2. Vanishing and Exploding Gradients

Figure S2 depicts the behavior of gradients in an $L = 100$ layer deep random CNN as σ_w^2 is varied across the phase boundary, from the ordered to the chaotic phase. In this case, the input size $n = 10$, kernel size $2k + 1 = 3$, and number of channels $c = 2000$. The input was synthetic and generated i.i.d from normal distributions with a random spatial covariance matrix. The y-axis plots the squared norm of the gradient with respect to the weights in layer ℓ, $||\nabla_{\boldsymbol{W}^\ell} h^L||_2^2$, with $\ell = 0, ..., 100$, relative to the last layer gradient, $||\nabla_{\boldsymbol{W}^L} h^L||_2^2$. The bias variance is fixed, $\sigma_b^2 = 0.05$, while the weight variance is swept from $\sigma_w^2 = 1.0$ (blue curves), where gradients vanish exponentially as a function of layer distance $L - \ell$ from the output, to $\sigma_w^2 = 4.25$ (red curves), where gradients explode exponentially. The mean field theory prediction (overlaid in dashed green) gives excellent agreement with the empirical result.

E. Distribution of singular values of weight matrices

The end-to-end Jacobian $\boldsymbol{J}$ depends on the matrix of weights $\boldsymbol{W}^l$, and the singular value distribution of the latter plays a key role, as discussed in the main text. Figure S3 compares the singular value distribution of weight matrices $\boldsymbol{W}^l$ in the convolutional vs. fully-connected setting. In more detail, $\boldsymbol{W}^l$ in the convolutional case can be considered an $n \times n$ circulant tiling of $c \times c$ dense blocks, where each matrix element is generated i.i.d. from $\mathcal{N}(0, 1/(c(2k + 1)))$. For fixed $n = 26$ and $2k + 1 = 5$ we compute the singular value distribution, as c increases, for single draws. This is compared to the distribution for the weight matrix in the fully-connected setting, obtained from a dense $nc \times nc$ matrix $\boldsymbol{W}^l$ whose entries are drawn i.i.d from $\mathcal{N}(0, 1/(nc))$. We empirically find the agreement between the two improves as the channel number increases, suggesting that the random matrix theory analysis of (Pennington et al., 2017; 2018) carries over to the convolutional setting.

F. Multiple depth scales in signal propagation

Figure S4 empirically demonstrates the existence of multiple depth scales, as discussed in Section 2.1.5. We consider an ensemble of random CNNs and compute the average covariance matrix $\boldsymbol{\Sigma}^l$ as a function of depth. We consider

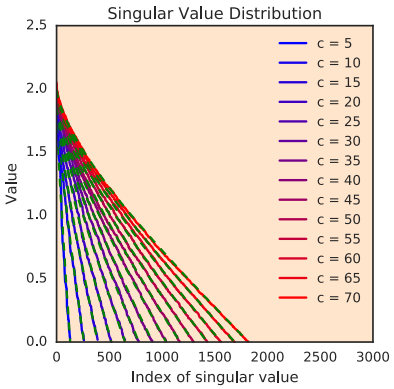

Figure S3. A comparison of the singular value distribution of $\boldsymbol{W}^l$ for a block-circulant matrix of shape $nc \times nc$ with i.i.d entries (solid) against that of an $nc \times nc$ dense matrix (overlaid in dashed green), for fixed spatial width $n = 26$ and kernel size $2k + 1 = 5$. While there are discrepancies for very small c that are not visible on this scale, there is good agreement as the channel size c increases (blue to red curves).

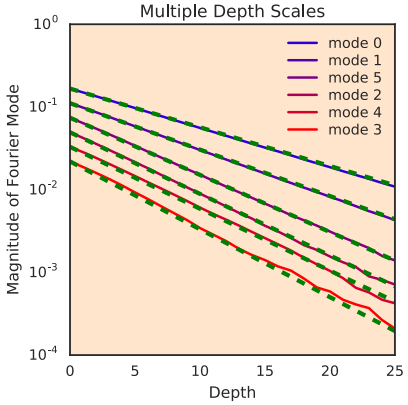

Figure S4. The existence of multiple depth scales in random CNNs, with a comparison between empirical results (colored) and theoretical predictions (dashed green). The depth scales are reflected in the differing slopes of the curves, with the zero frequency mode decaying most slowly. See Section F for a detailed description of the experiment.

networks with erf nonliearities with $\sigma_w = \frac{3}{2}$ and $\sigma_b = \frac{1}{2}$ applied to 1D images of size $n = 10$. The initial data covariance $\boldsymbol{\Sigma}^0$ is chosen so that $\boldsymbol{\epsilon}^0 = \boldsymbol{\Sigma}^* - \boldsymbol{\Sigma}^0$ is small and has an off-diagonal structure. In particular, all entries of $\boldsymbol{\epsilon}^0$ except the first cyclic diagonal entries are taken to be zero, and that diagonal has Fourier transform given by $-\frac{1}{6}[1, \frac{2}{3}, (\frac{2}{3})^3, (\frac{2}{3})^5, (\frac{2}{3})^4, (\frac{2}{3})^2, (\frac{2}{3})^4, (\frac{2}{3})^5, (\frac{2}{3})^3, \frac{2}{3}]$. We used a spatially non-uniform kernel of size $2k + 1 = 3$, with weights $v = [0.025, 0.950, 0.025]$. We then averaged $\boldsymbol{\Sigma}^l$ over this ensemble of networks to construct $\boldsymbol{\epsilon}^l$. By decomposing the vector of first cyclic diagonal entries into Fourier modes, we can observe how the signal decays differently along different modes. Figure S4 plots the absolute value of the coefficient of the Fourier decomposition as a function of depth. Our mean field theory predictions for the different depth scales are in excellent agreement with the empirical simulations.