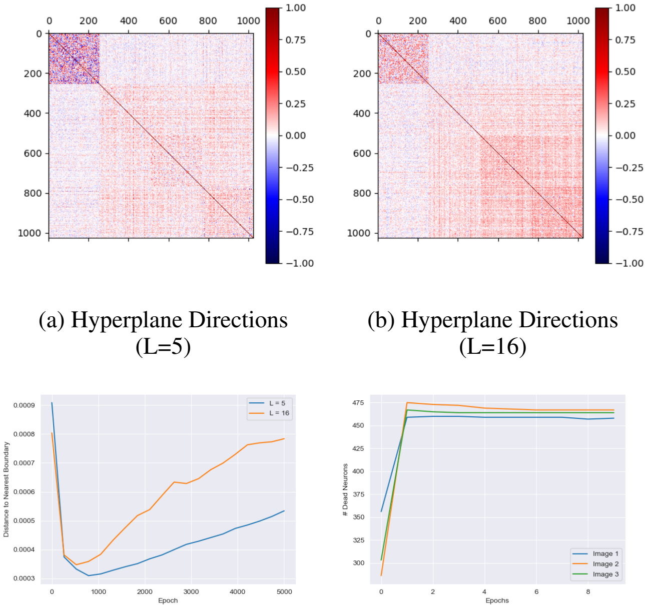
Appendix

1 Natural Images

In this section, we would like to test our ideas in a slightly more practical setting by recreating certain experiments on natural images. We also test with a different architecture as well, using a four layer MLP with 256 hidden neurons. In the main text, we used a random image to denote a high frequency 2D signal in order to clearly demonstrate the limitations of dense coordinates. In natural images or signals, there may be low frequency regions or constant regions as opposed to solely high frequencies, hence the dynamics may be slightly different. However, as positional encoding still achieves superior performance, and spectral bias is still present in this setting, the overall relationship between expressive capacity and confusion should still be present.

In Figure 1 (a) and (b), we show the correlation of hyperplane directions after training on a natural image to demonstrate the same behavior found in the main text. We also find that the deeper layers become progressively correlated with higher frequency encodings as well. In Figure 1 (c), we plot the boundary distances of training points which displays wider regions with higher frequency encodings. In Figure 1 (d), we can see the dying ReLU problem still occurs due to the density of coordinates, and is relatively the same for each image.

In Figure 2, we show that the relationship between expressive capacity and confusion applies for natural signals as well. For both methods, there is more confusion locally compared to globally. This is still a result of an increase in the hamming distance for globally sampled inputs in relation to locally sampled inputs, and positional encoding reduces the severity of this confusion across the signal due to its enhanced expressive power. Since the images now contain regions of constant or slowly changing target values, we see slight differences in the confusion densities. As training progresses, the gradient directions for globally sampled coordinates become increasingly orthogonal, a result of their rise in hamming distance during training. As these only correspond to the low frequency components, they do not indicate a reduction of spectral bias with coordinates, but rather that the network can better speed convergence to these components on natural signals using a larger architecture. However, the local densities are still the widest, meaning the high frequency components converge very slowly. Additionally, the gradient directions for positional encoding tend to lean towards positively correlated, a result of their positively correlated hyperplane directions. The effects of this are also shown in the large dip in mean hamming distance. For further insight on this behavior in the non 2D setting, we show similar results for a 1D signal in Figure 4. Overall, the same principles apply to various signals, although the severity of confusion and its dynamics depend on the amount of high frequency components within the signal, in addition to the architecture used.



(a) Hyperplane Directions (L=5)

(b) Hyperplane Directions (L=16)

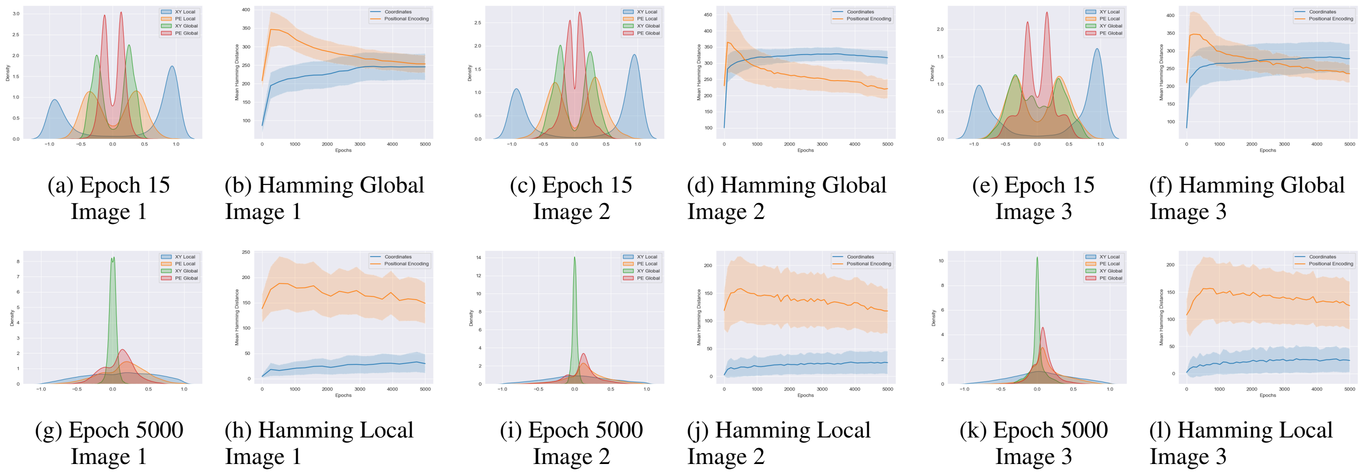
(c) Distance to Boundary

(d) Dead Coordinate Neurons

Figure 1: Positively correlated hyperplanes and wider regions demonstrated on a natural image and 4 layer MLP. Higher frequency encodings behave the same for natural signals, and we can see how their directions become higher correlated with depth. For (c), we use mini-batches as opposed to batch gradient descent shown in the main text. In (d), we show that coordinates still induce dead neurons during training.

2 Restricted Positional Encoding

In the main text, we mentioned how scaling the frequency or dimensionality of the encoding alone does not result in more



(a) Epoch 15 Image 1

(b) Hamming Global Image 1

(c) Epoch 15 Image 2

(d) Hamming Global Image 2

(e) Epoch 15 Image 3

(f) Hamming Global Image 3

(g) Epoch 5000 Image 1

(h) Hamming Local Image 1

(i) Epoch 5000 Image 2

(j) Hamming Local Image 2

(k) Epoch 5000 Image 3

(l) Hamming Local Image 3

Figure 2: Hamming distance and confusion densities for 3 different natural images. We used 100 20x20 local regions and 25,000 pairs of distant inputs to compare, with a high encoding frequency of L=16. We see similar trends to those presented in the main text. One of these similarities is the larger hamming distance across the signal when using positional encoding, which appears to dip as the hyperplanes become more positively correlated. Additionally, both methods contain less global confusion and more local confusion which aligns with their hamming distances, although positional encoding reduces the gap between them. The coordinate based networks are able to reduce confusion globally better than in the main text, most likely the result of a lower frequency signal and larger network. However, convergence is still greatly hindered for the high frequency components, which explains the observed spectral bias.

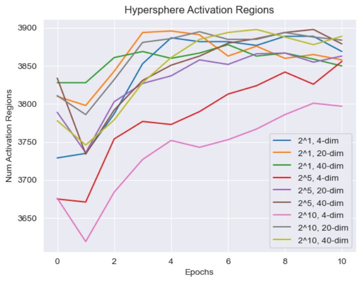


Figure 3: Growth of activation regions with different length and frequency encodings. As the frequency is the same across each dimension, the input is of a higher density and expressive power is similar to that of coordinates as shown in the main text.

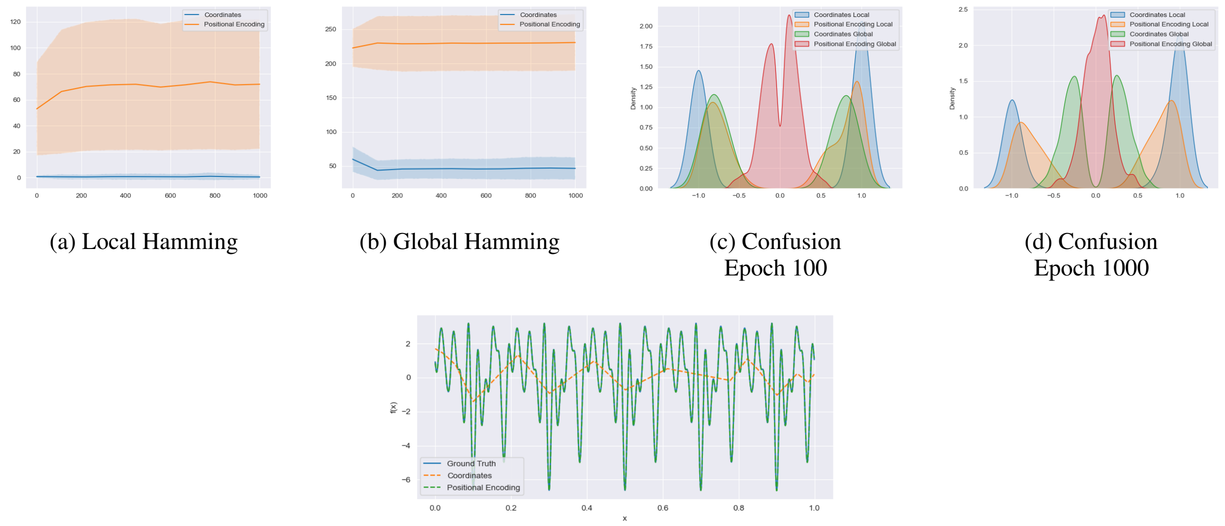
expressive capacity utilized by the model. If we restrict the encoding to use the same frequency along each dimension, then the density imposed by coordinates is the same in a higher dimensional space. We mention this as it demonstrates how a proper encoding of coordinates is needed in order to truly alleviate their density and generate a lower frequency function, as simply mapping them to higher dimensions will not always result in more expressive power. We display this in Figure 3, where we test different frequencies as well as higher dimensions, and count the total number of regions utilized by the network during training on the 64x64 random image. For each method, the number of regions is lower than the number of examples in the dataset.

3 Depth and Width

In this section, we continue to examine the hamming distance on the random 2D image, however we focus on the scaling of the network and how this effects the relative distinctiveness of linear functions. By relative, we mean the ratio between the hamming distance and the number of neurons, which can help determine how unique each linear function is relative to the size of the network. Since many coordinates will lie in similar regions in the first layer, it will require more layers to separate each input to a unique region. This also means the activation patterns should become more distinct in later layers as opposed to earlier layers. Due to the effects of positional encoding discussed in the main text, the inputs should be properly dispersed across the regions in the first layer, which can easily propagate throughout the following layers.

In Figure 6, we display this scaling using different sized networks of two, four, and six layers, containing 128, 256, and 512 hidden neurons respectively. We plot both the ratios and the overall hamming distance of the network to evaluate. We can see in local regions, the ratio is much higher for networks with positional encoding, meaning the linear functions the network utilizes are increasingly distinct relative to the total number of neurons. With coordinates however, the ratio does not change much, which displays the difficulties that arise from high density inputs and the sheer size of the network that will be needed to overcome it. One interesting aspect of these plots is that the ratio continues to decrease as the size of the network grows larger, meaning many of the available regions are not utilized. However, while the ratio decreases with scale, the true hamming distance increases as expected.

In Figure 6 (e) and (f), we display the confusion densities at the end of training. We can see an interesting relationship between the ratio of hamming distance and true hamming distance when reducing confusion. For example, take the six layer coordinate network in comparison to the two layer encoding network. The true hamming distance for the six layer coor-



(a) Local Hamming

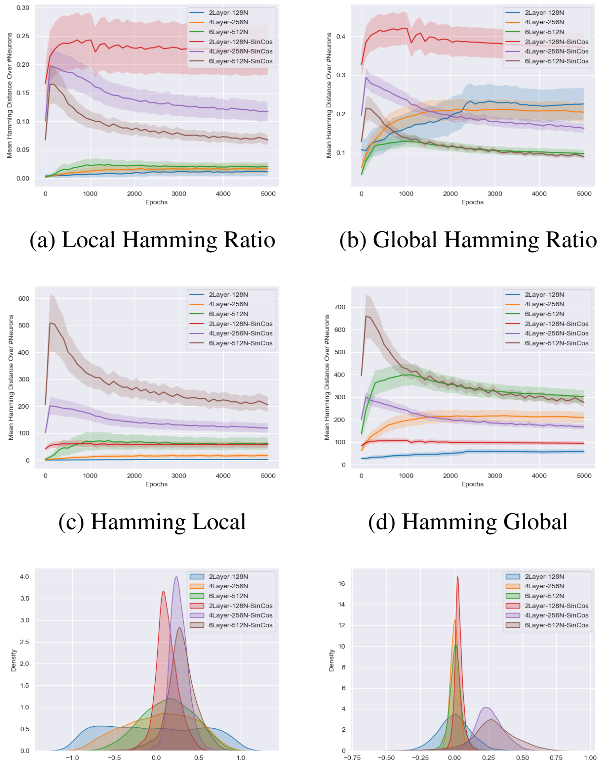
(b) Global Hamming

(c) Confusion Epoch 100

(d) Confusion Epoch 1000

(e) Results (Epoch 1000)

Figure 4: Standard results on a 1D signal using a 2 layer MLP with 256 neurons, and an L value of 5 for the encoding. The same general principles apply in this setting as well.



(a) Local Hamming Ratio

(b) Global Hamming Ratio

(c) Hamming Local

(d) Hamming Global

(e) Confusion Local

(f) Confusion Global

Figure 5: Ratio between hamming distance and true hamming distance on the random 2D image. While the true hamming distance increases with the total number of neurons, there is a dip in their ratio. As this ratio is already restricted with coordinate based inputs, the benefits of increased depth and width are minimal.

dinate network is much larger, however its relative hamming distance is restricted which can still result in higher amounts of confusion due to the utilization of similar linear functions. While the global confusion densities (f) between the two layer encoding network and six layer coordinate network are slowly becoming more similar, this demonstrates the severity of the limitations imposed by the density of coordinates, and the size of the network that will be needed in order to obtain similar performance to positional encoding.