
YAMATANI ACTIVATION: EDGE HOMOGENEOUS RESPONSE SUPER RESOLUTION NEURAL NETWORK

A PREPRINT

Takuma Yoshimura*
poco-apoco Networks Co. Ltd.

February 19, 2020

ABSTRACT

In this research, I propose a two-variable activation function "*Yamatani*" that satisfies the first-degree homogeneity, and realize a super-resolution convolutional neural network that is independent of the dynamic range and symmetrical about the luminance inversion.

Keywords Super Resolution · Convolutional Neural Network · Homogeneous Response Activation

1 Introduction

The performance of single-image super-resolution, which enlarges an image while maintaining sharpness, has been significantly improved since the introduction of Convolutional Neural Networks(hereinafter called CNN).

On the other hand, CNN-based super-resolution does not produce the same result for images with different dynamic ranges due to bias parameters, and there is no guarantee that the same sharpened pattern will be output at edges with different luminance differences from the surroundings. Furthermore, when a non-linear activation function such as Relu or leakyRelu is used, there is a point that luminance inversion symmetry is not guaranteed.

This study proposes an activation function "*Yamatani*" that satisfies homogeneity with two variables as inputs. Yamatani activation is derived from the origami "*Yamaori*" (mountain fold) "*Taniori*" (valley fold). By using this activation function, it is possible to realize super-resolution that does not depend on a dynamic range in which a satisfies homogeneity residual responds to a luminance difference between an edge and surrounding pixels.

2 Homogeneity, Additivity and Linearity

The section supplements about homogeneity, additivity and linearity.

2.1 Definition

That a function f satisfies the n -degree homogeneity means that the following equation is satisfied for the input vector \mathbf{v} .

$$\forall \mathbf{v} \in \mathbb{R}^N, k \in \mathbb{R}[k^n \cdot f(\mathbf{v}) = f(k \cdot \mathbf{v})]. \quad (1)$$

Also, that a function f satisfies the additivity means that the following equation is satisfied for the input vector \mathbf{u}, \mathbf{v} .

$$\forall \mathbf{u}, \mathbf{v} \in \mathbb{R}^N [f(\mathbf{u}) + f(\mathbf{v}) = f(\mathbf{u} + \mathbf{v})]. \quad (2)$$

Further, that the function f is linear means that it satisfies first-degree homogeneity and additivity.

*github: <https://github.com/tk-yoshimura/>

2.2 Homogeneous Response Activation

The response of a CNN cannot be approximated to any function without using a nonlinear activation function. However, it is nonlinear if it satisfies homogeneity and does not satisfy additivity. In the univariate activation function, the first-degree homogeneous function is only the proportional function, which is equivalent to using a linear function as the activation function. That is, with an activation function of one variable, a CNN that does not depend on the dynamic range and satisfies the inversion symmetry cannot be constructed. Therefore, I introduce a two-variable activation function that satisfies first-degree homogeneity.

3 Yamatani activation

The section describes the definition and properties of yamatani activation.

3.1 Definition

Yamatani activation is defined by the following equation using two variables x, y and constant $s \geq 0$ as inputs. This activation a curved surface as shown in Figure 1.

$$Yamatani(x, y; s) := s \cdot (x + y) + \begin{cases} \min(x, y) & (x > 0, y > 0), \\ \max(x, y) & (x < 0, y < 0), \\ 0 & (\text{otherwise}). \end{cases} \quad (3)$$

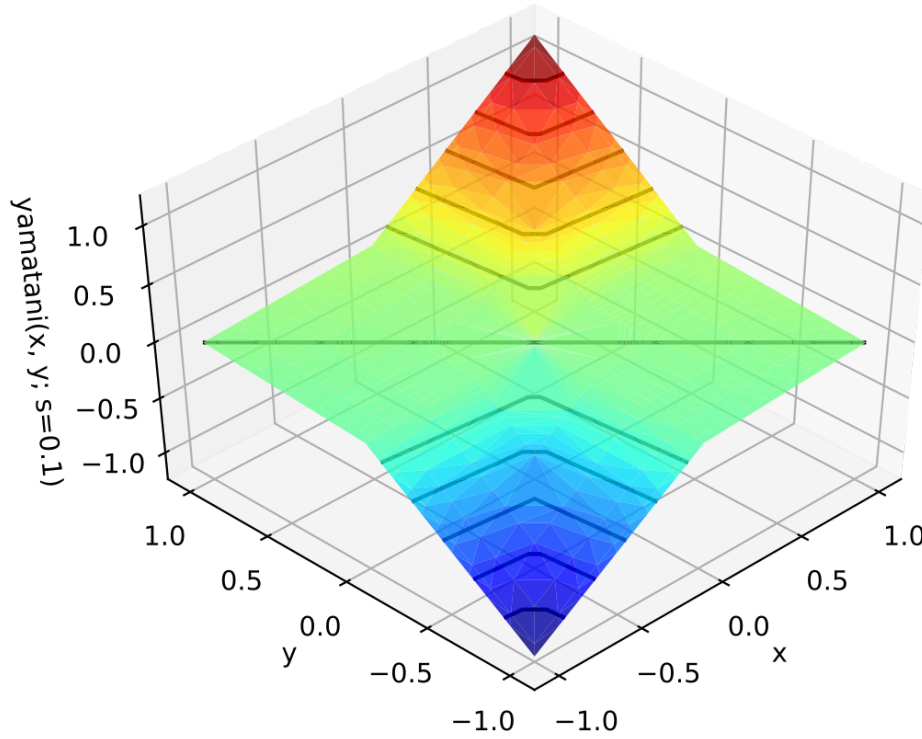


Figure 1: Yamatani Surface

This function satisfies the first-degree homogeneity for the input x, y .

$$\forall x, y, k \in \mathbb{R} [k \cdot Yamatani(x, y; s) = Yamatani(k \cdot x, k \cdot y; s)]. \quad (4)$$

The convolutional layer with yamatani activation as an activation function is defined as follows. In the expression, \mathbf{v} is input vector, $\mathbf{W}_1, \mathbf{W}_2$ is weight parameters of the convolution function, and $*$ is a convolution operator. Bias parameter is NOT used.

$$YamataniConv(\mathbf{v}; \mathbf{W}_1, \mathbf{W}_2) := Yamatani(\mathbf{v} * \mathbf{W}_1, \mathbf{v} * \mathbf{W}_2; s). \quad (5)$$

The convolutional layer satisfies the first-degree homogeneity for the input vector v .

$$\forall v \in \mathbb{R}^N, k \in \mathbb{R} [k \cdot \text{YamataniConv}(v; \mathbf{W}_1, \mathbf{W}_2) = \text{YamataniConv}(k \cdot v; \mathbf{W}_1, \mathbf{W}_2)]. \quad (6)$$

Obviously, even when used as a residual connection, it satisfies homogeneity.

$$\text{ResYamataniConv}(v; \mathbf{W}_1, \mathbf{W}_2) := v + \text{YamataniConv}(v; \mathbf{W}_1, \mathbf{W}_2). \quad (7)$$

3.2 Propertie

When the convolutional layer using yamatani as the activation function is applied recursively, the following surface is obtained, which can approximate any first-degree homogeneous multivariable function.

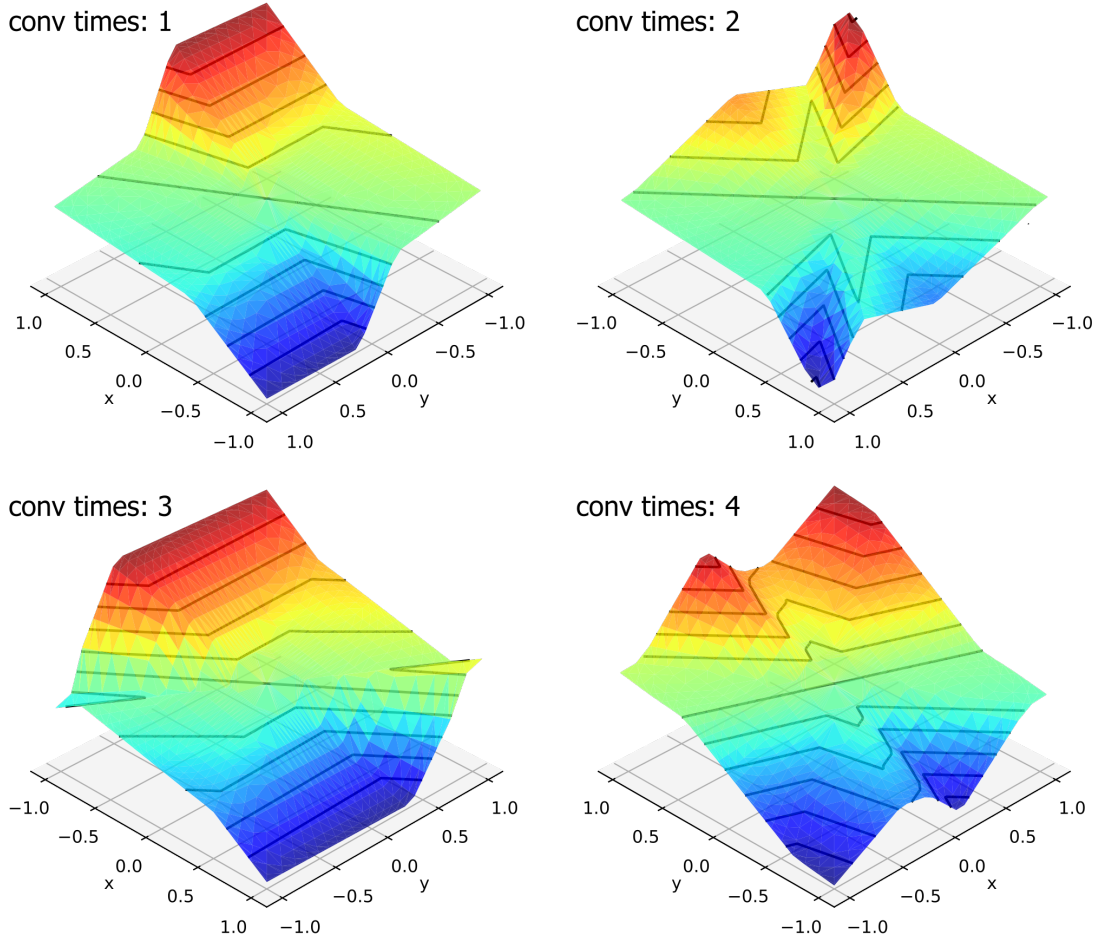


Figure 2: Yamatani Convolution Complexity

4 Super-resolution algorithm and symmetry to be satisfied

The symmetry that the super-resolution algorithm should satisfy includes spatial symmetry and luminance symmetry.

Spatial symmetry refers to the fact that when the image is rotated or inverted and then super-resolution is performed and then the image is restored, the same result is obtained as an image that has not been rotated or inverted(Figure 3). A conventional method for satisfying the spatial symmetry described in the previous section is self-ensemble. This is a method of super-resolution of 8 rotated and inverted images and taking the average [1].

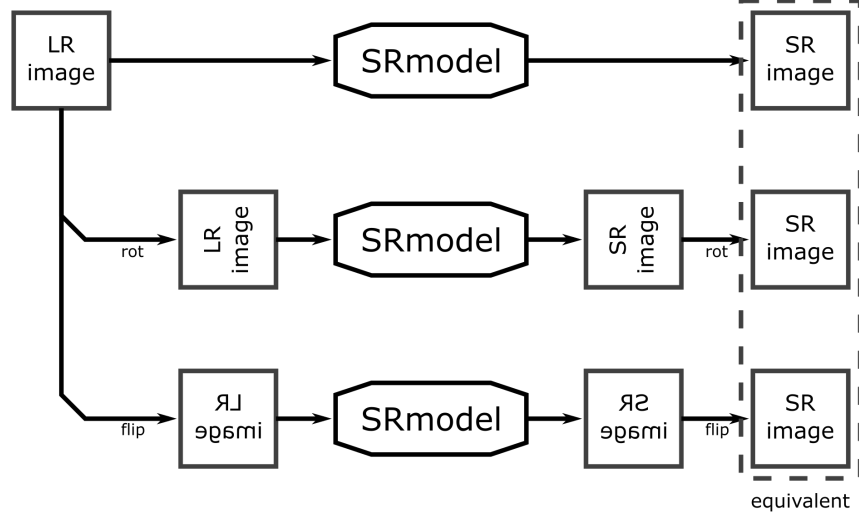


Figure 3: Spatial Symmetry

Luminance symmetry means that an image obtained by applying a linear transformation to each luminance value of an image and performing super-resolution and then inversely transforming the luminance value is equivalent to an image not subjected to a linear transformation(Figure 4).

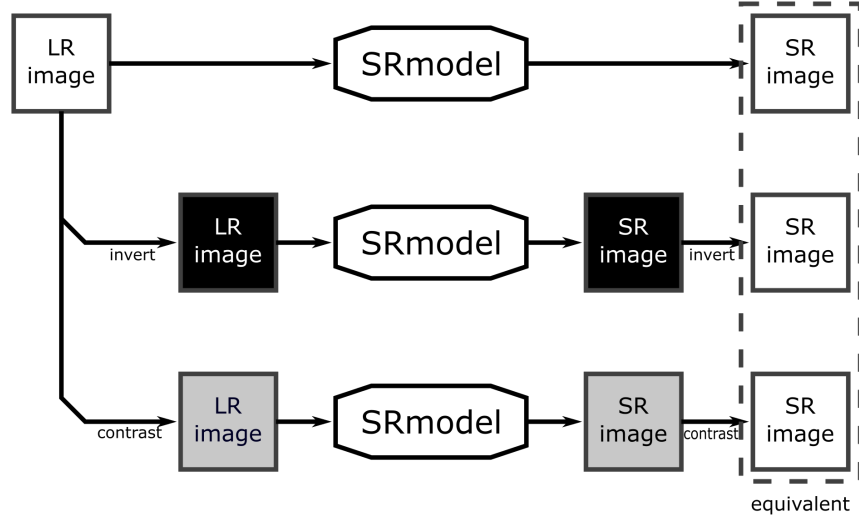


Figure 4: Luminance Symmetry

5 Yamatani-based Homogeneous Super Resolution Network

The following is proposed as a super-resolution model using yamatani activation. This model is referenced to DRRN [2]. Yamatani-based Homogeneous Super Resolution Network(hereinafter called YHSRN) consists of three blocks, entry block, residual block, and terminal block. In the convolution layer of the entry block, the sum of the spatial parameter directions is limited to 0, and only the difference value of the image is output. Pixel Shuffler is referenced to [3].

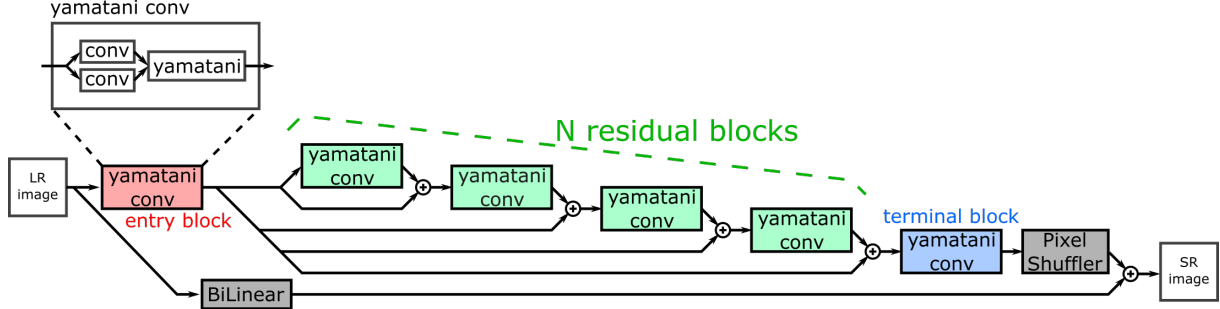


Figure 5: Yamatani-based Homogeneous Super Resolution Network

The following table shows the hyperparameters related to the structure of YHSRN. Generally, I is 1 of the Y components in the YUV color space, or 3 of all the components in the RGB or YUV color spaces. The number of output channels of the terminal block is IS^2 .

category	parameter	symbol
entry block	In Channels	I
	Kernel Size	K_E
residual block	Channels	C
	Blocks	N
	Kernel Size	K_R
terminal block	Kernel Size	K_T
-	Enlargement Ratio	S

Table 1: Yamatani-based Homogeneous Super Resolution Network Hyperparameters

Bilinear was interpolated by the following equation so that the interpolation coefficient did not depend on the pixel position. In the expression, x, y is pixel index, $S_{x,y}$ is source image values, and $D_{x,y}$ is destination image value. Reference to a position outside the image is reinterpreted as the nearest position in the image. Here, the case where the enlargement ratio S is 2 will be described.

$$\begin{bmatrix} D_{2x,2y} \\ D_{2x+1,2y} \\ D_{2x,2y+1} \\ D_{2x+1,2y+1} \end{bmatrix} = \frac{1}{9} \begin{bmatrix} 4S_{x,y} + 2S_{x-1,y} + 2S_{x,y-1} + S_{x-1,y-1} \\ 4S_{x,y} + 2S_{x+1,y} + 2S_{x,y-1} + S_{x+1,y-1} \\ 4S_{x,y} + 2S_{x-1,y} + 2S_{x,y+1} + S_{x-1,y+1} \\ 4S_{x,y} + 2S_{x+1,y} + 2S_{x,y+1} + S_{x+1,y+1} \end{bmatrix}. \quad (8)$$

6 Functional verification

To check the homogeneous response of YHSRN, training is performed with one image that has not been subjected to linear transformation, and then the image that has been subjected to linear transformation is inferred to verify whether super-resolution has been properly performed (Figure 6). Here, $I = 3$, $K_E = 7$, $C = 32$, $N = 3$, $K_R = 3$, $K_T = 3$, $S = 2$ as hyperparameters in Table 1. Edge padding is used as a padding method.

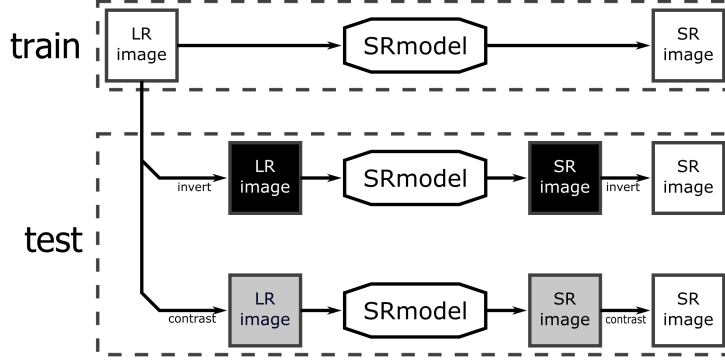


Figure 6: Functional verification

I use Lanna’s halved image as the verification image. The image before reduction is learned as a target image (Figure 7). The results are shown below (Figure 8). Then multiply the input image by -1 and the output image by -1. The same image was obtained and the luminance inversion symmetry was confirmed (Figure 9). Then multiply the input image by 0.5 and the output image by 2. In this case, the same image was obtained, and it was confirmed that the image did not depend on the dynamic range (Figure 10).



Figure 7: Training target



Figure 8: Training results



Figure 9: Luminance inversion symmetry



Figure 10: Independent of dynamic range

References

- [1] Radu Timofte, Rasmus Rothe, Luc Van Gool, Seven ways to improve example-based single image super resolution, CVPR, 2010.
- [2] Ying Tai, Jian Yang, and Xiaoming Liu, Image Super-Resolution via Deep Recursive Residual Network, CVPR, 2017.
- [3] Wenzhe Shi, Jose Caballer, Ferenc Huszar, Johannes Totz, Andrew P. Aitken, Rob Bishop, Daniel Rueckert, Zehan Wang, Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network, CVPR, 2016.