Supplementary Material: Pixel Recursive Super Resolution

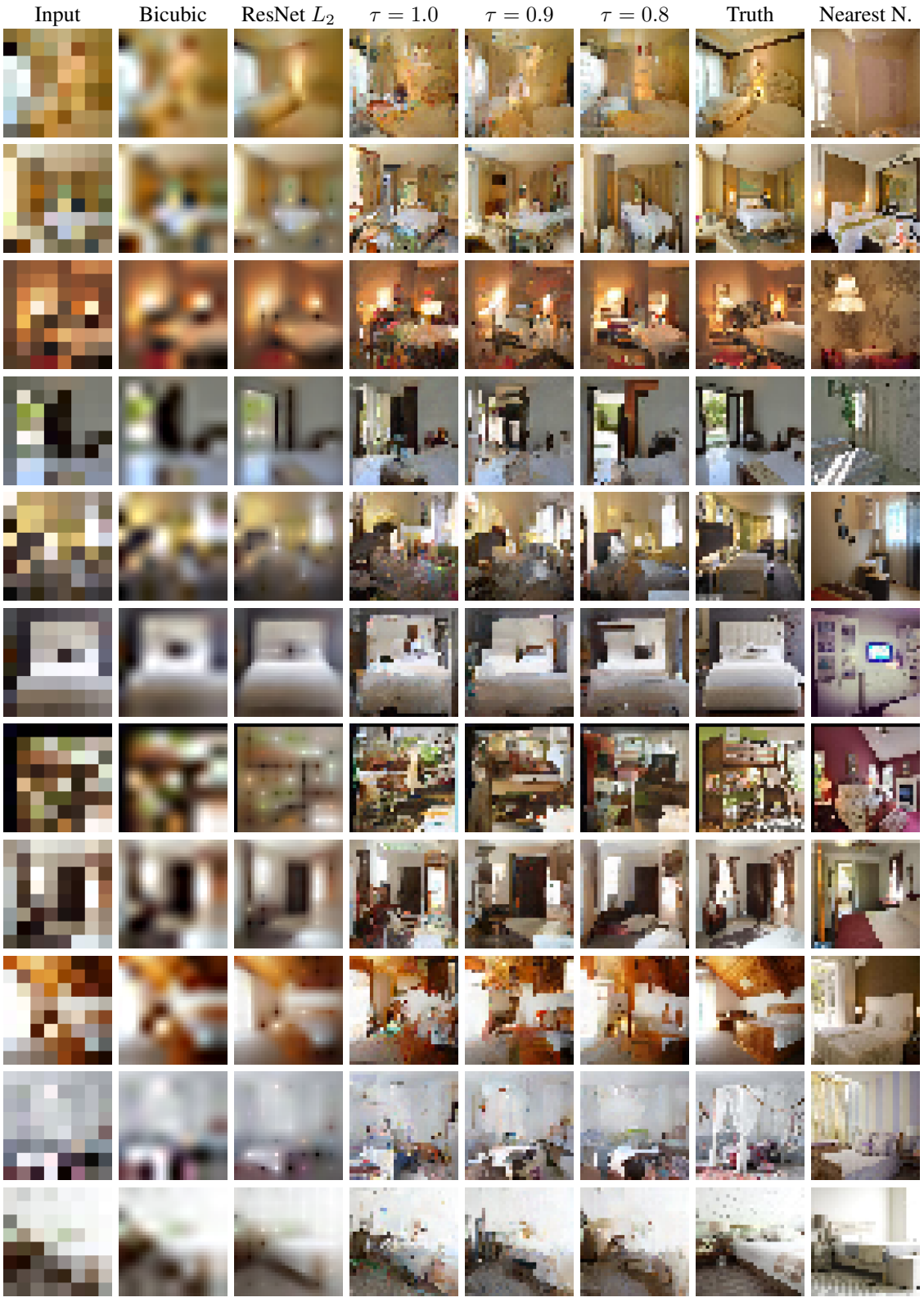
Ryan Dahl Mohammad Norouzi Jonathon Shlens
Google Brain
{rld,mnorouzi,shlens}@google.com

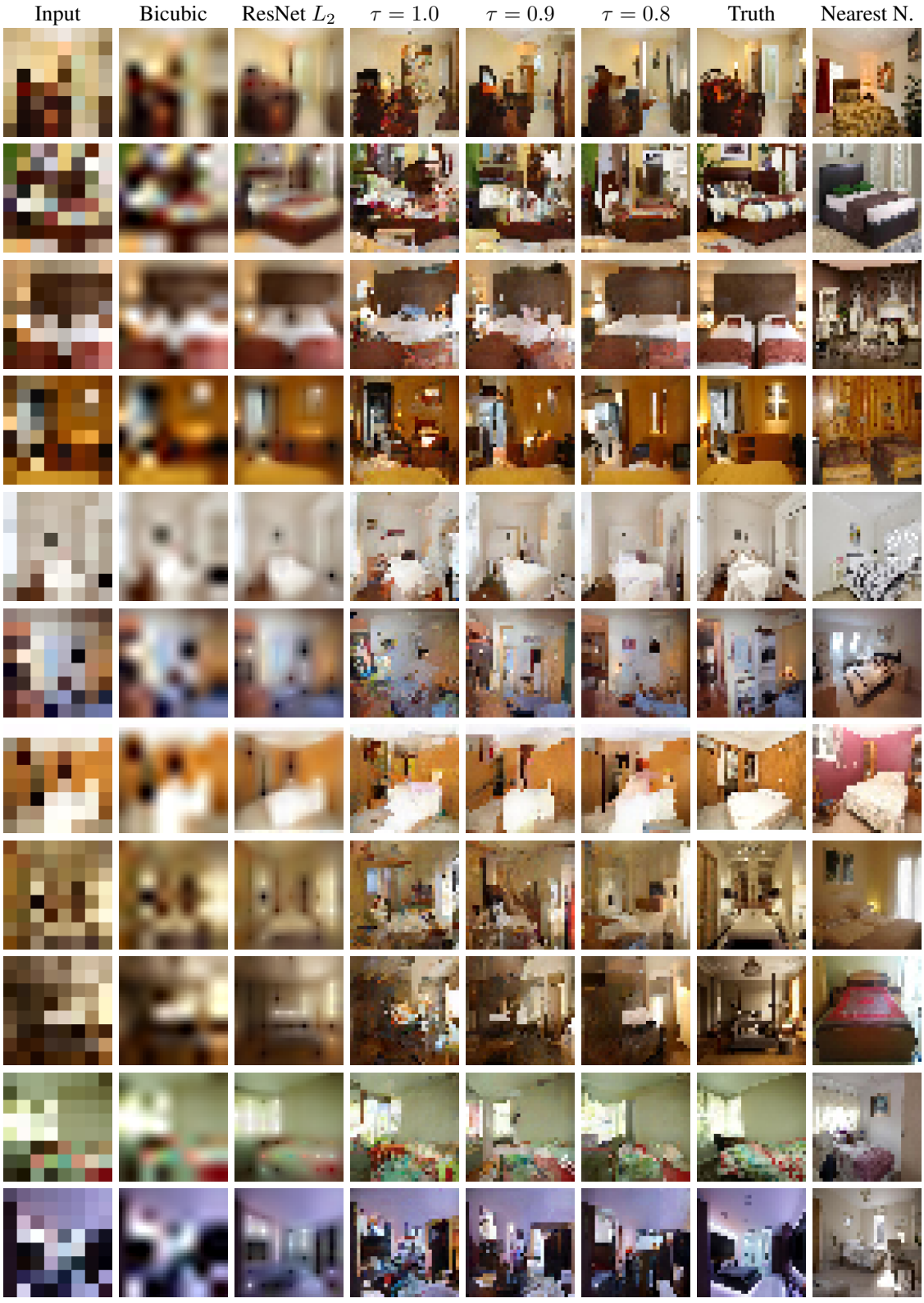
1. Hyperparameters for pixel recursive super resolution model.

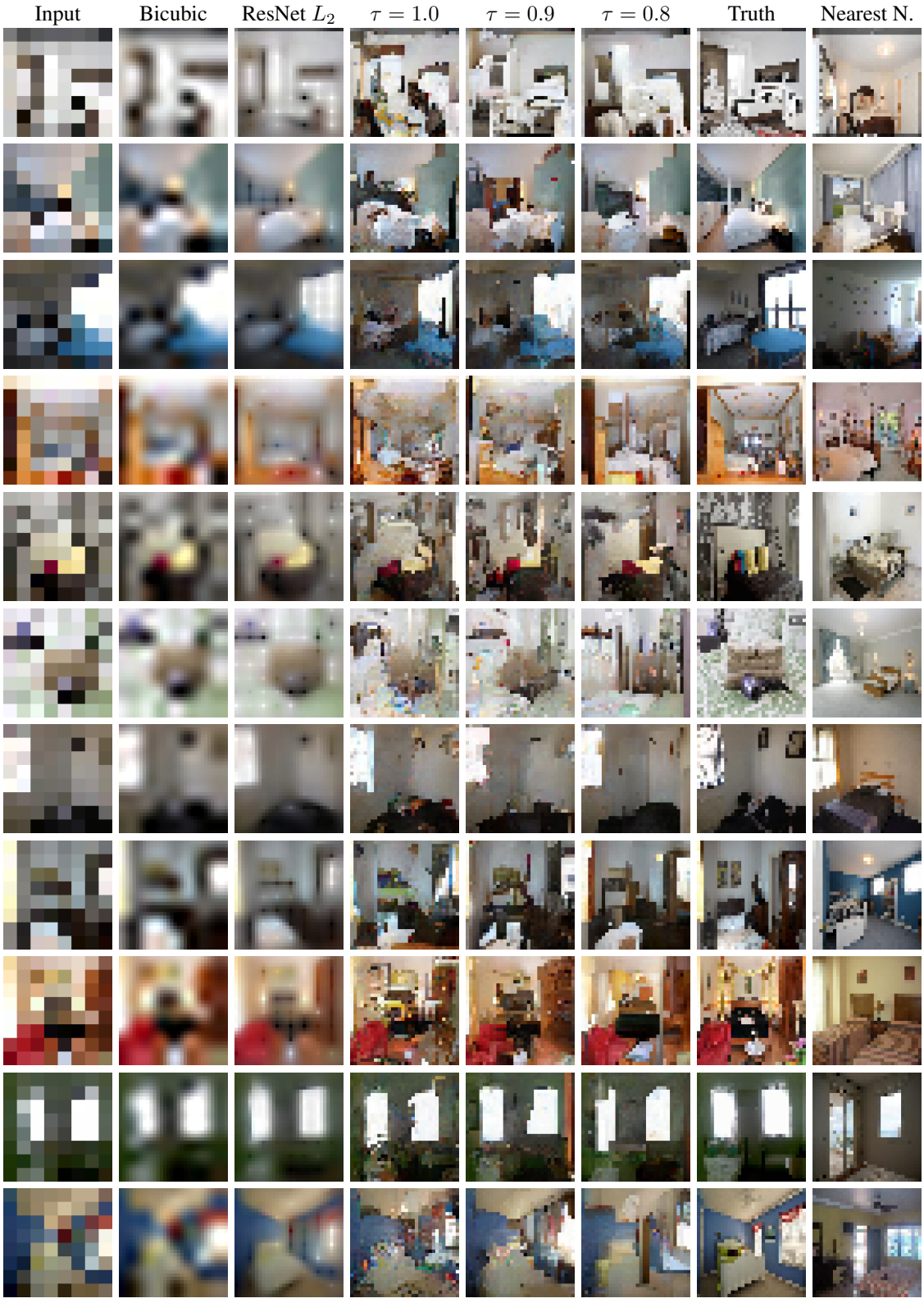
Operation	Kernel	Strides	Feature maps
Conditional network – $8 \times 8 \times 3$ input			
$B \times$ ResNet block	3×3	1	32
Transposed Convolution	3×3	2	32
$B \times$ ResNet block	3×3	1	32
Transposed Convolution	3×3	2	32
$B \times$ ResNet block	3×3	1	32
Convolution	1×1	1	$3 * 256$
PixelCNN network – $32 \times 32 \times 3$ input			
Masked Convolution	7×7	1	64
$20 \times$ Gated Convolution Blocks	5×5	1	64
Masked Convolution	1×1	1	1024
Masked Convolution	1×1	1	$3 * 256$
Optimizer	RMSProp (decay=0.95, momentum=0.9, epsilon=1e-8)		
Batch size	32		
Iterations	2,000,000 for Bedrooms, 200,000 for faces.		
Learning Rate	0.0004 and divide by 2 every 500000 steps.		
Weight, bias initialization	truncated normal (stddev=0.1), Constant(0)		

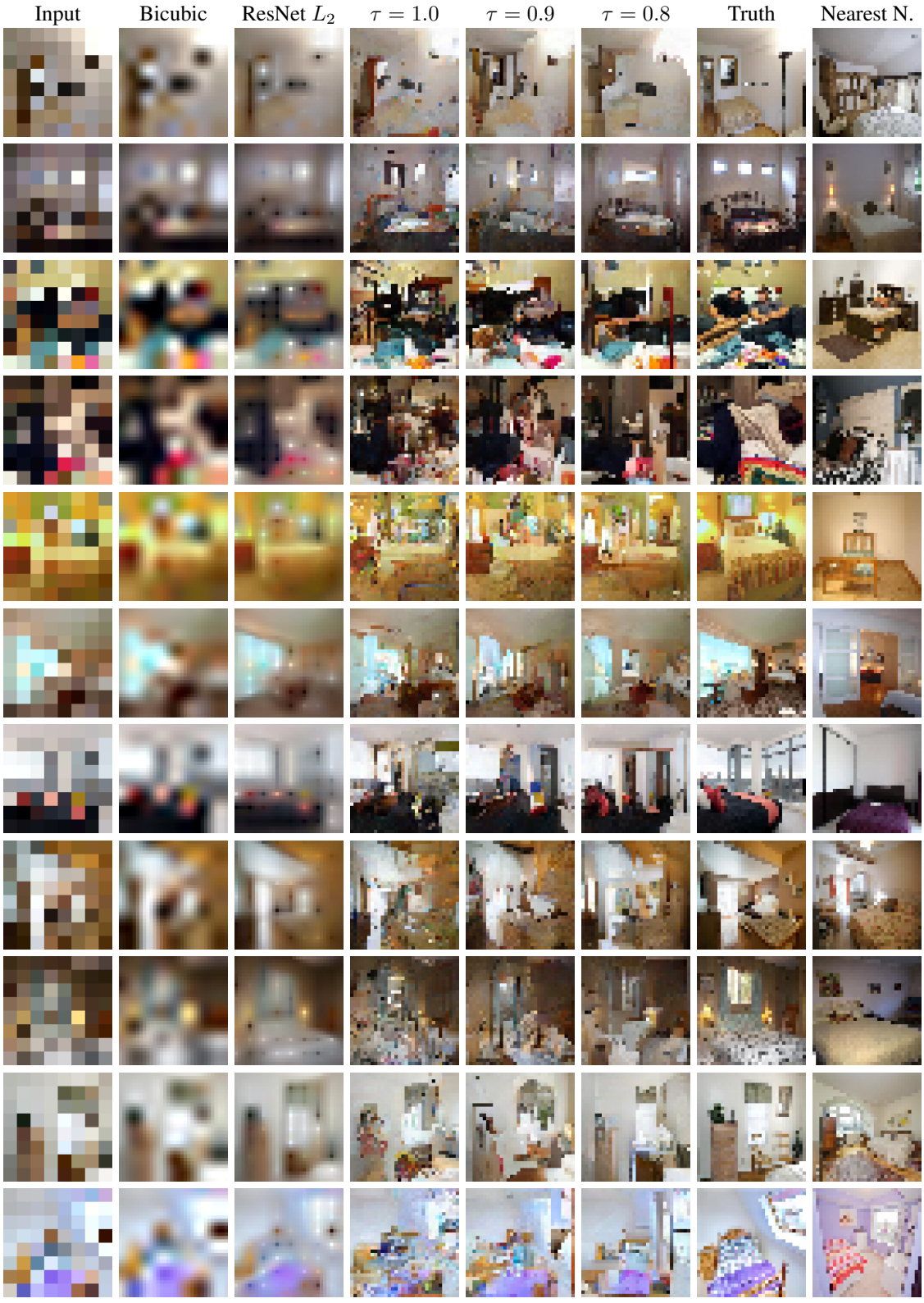
Table 1: Hyperparameters used for both datasets. For LSUN bedrooms $B = 10$ and for the cropped CelebA faces $B = 6$.

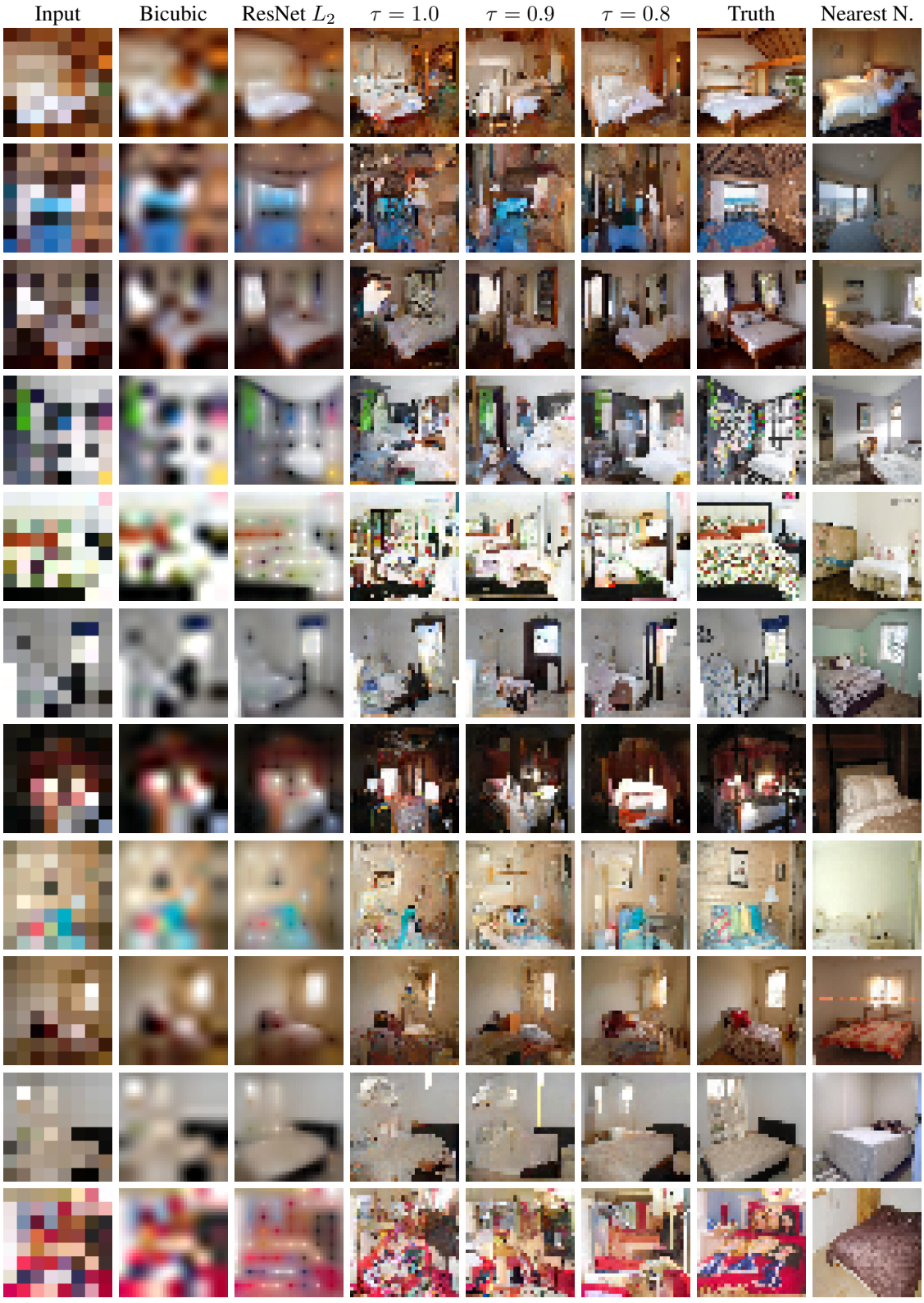
2. Samples from models trained on LSUN bedrooms



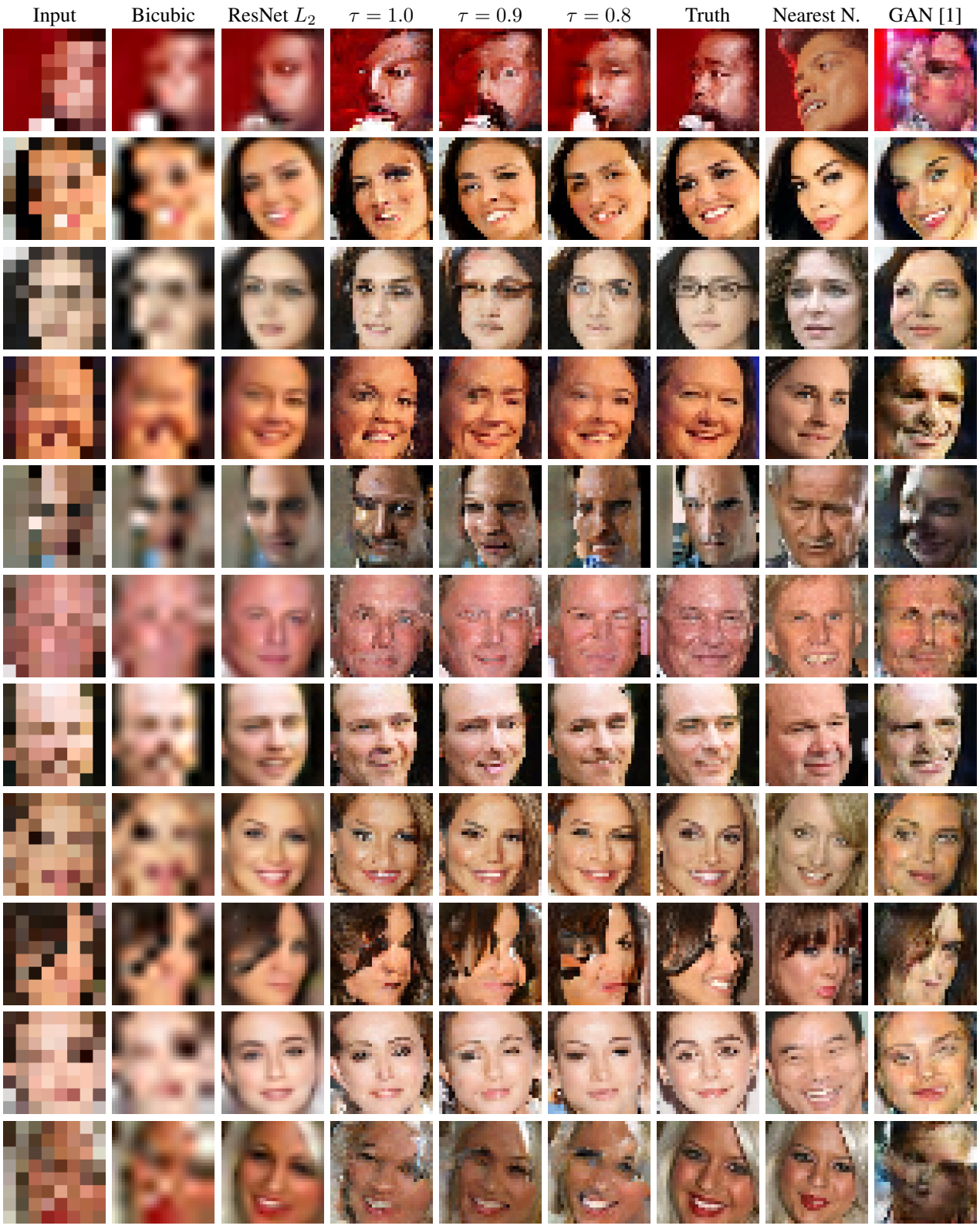

Input
Bicubic
ResNet L_2
$\tau = 1.0$
$\tau = 0.9$
$\tau = 0.8$
Truth
Nearest N.

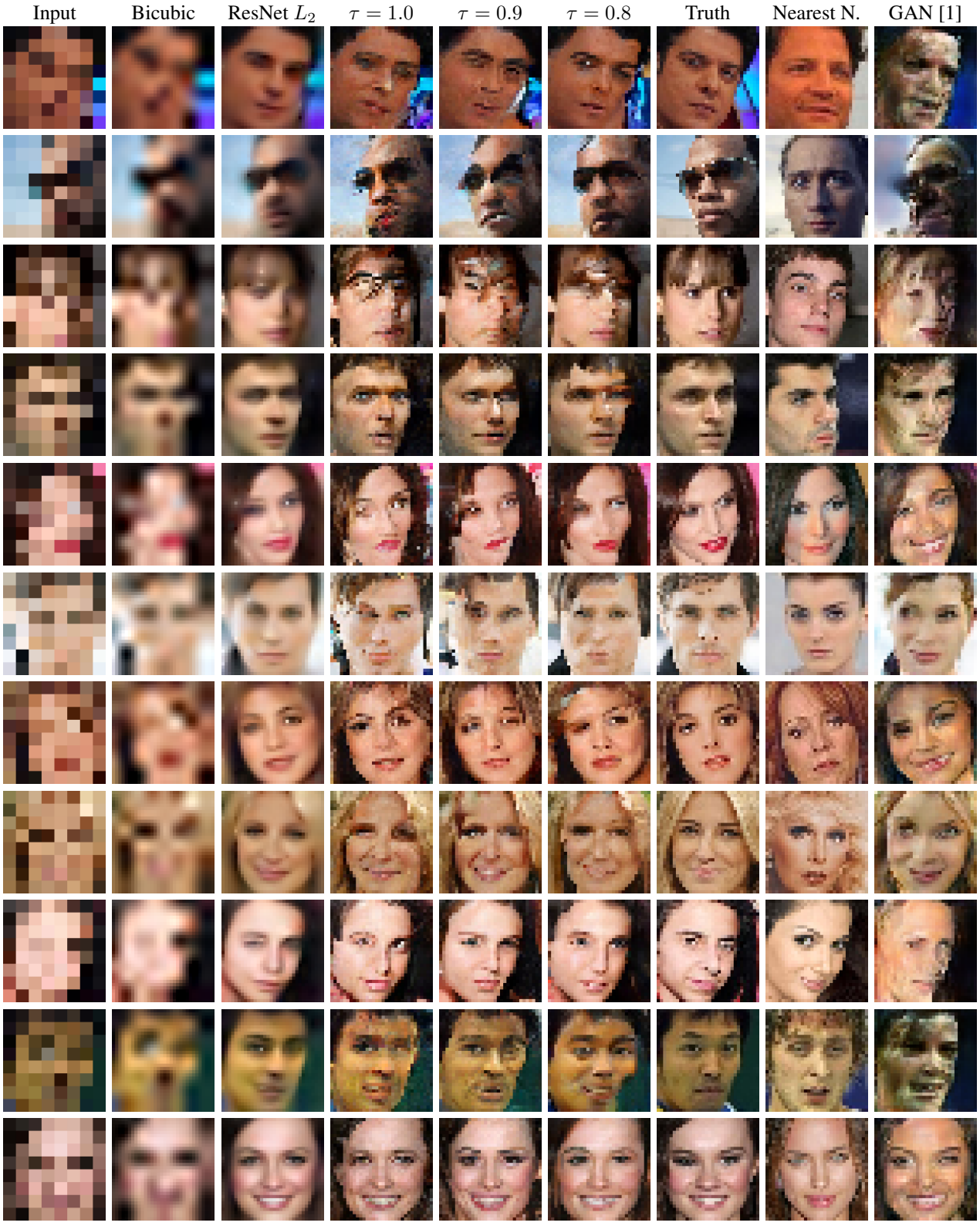

Input
Bicubic
ResNet L_2
$\tau = 1.0$
$\tau = 0.9$
$\tau = 0.8$
Truth
Nearest N.

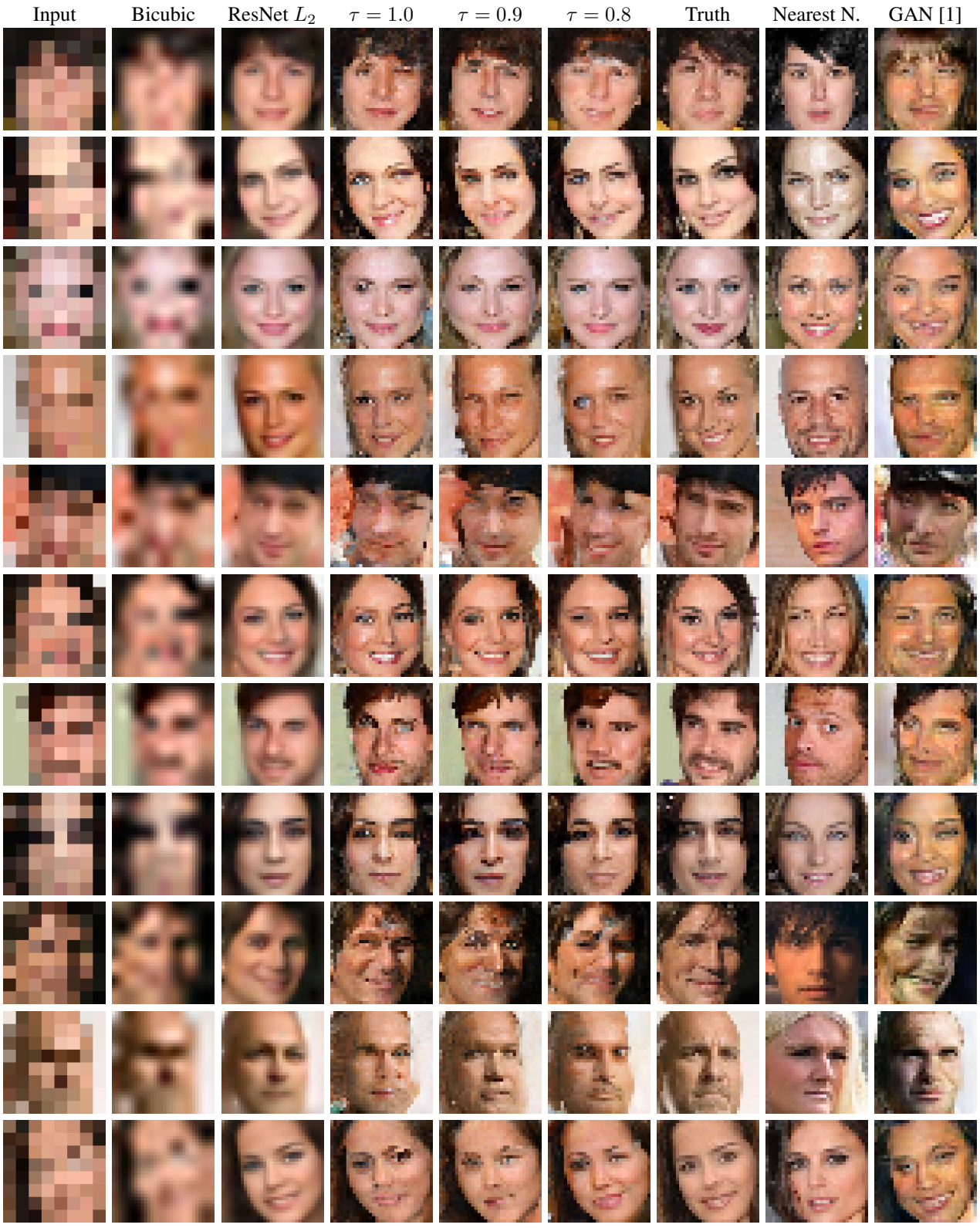

Input
Bicubic
ResNet L_2
$\tau = 1.0$
$\tau = 0.9$
$\tau = 0.8$
Truth
Nearest N.

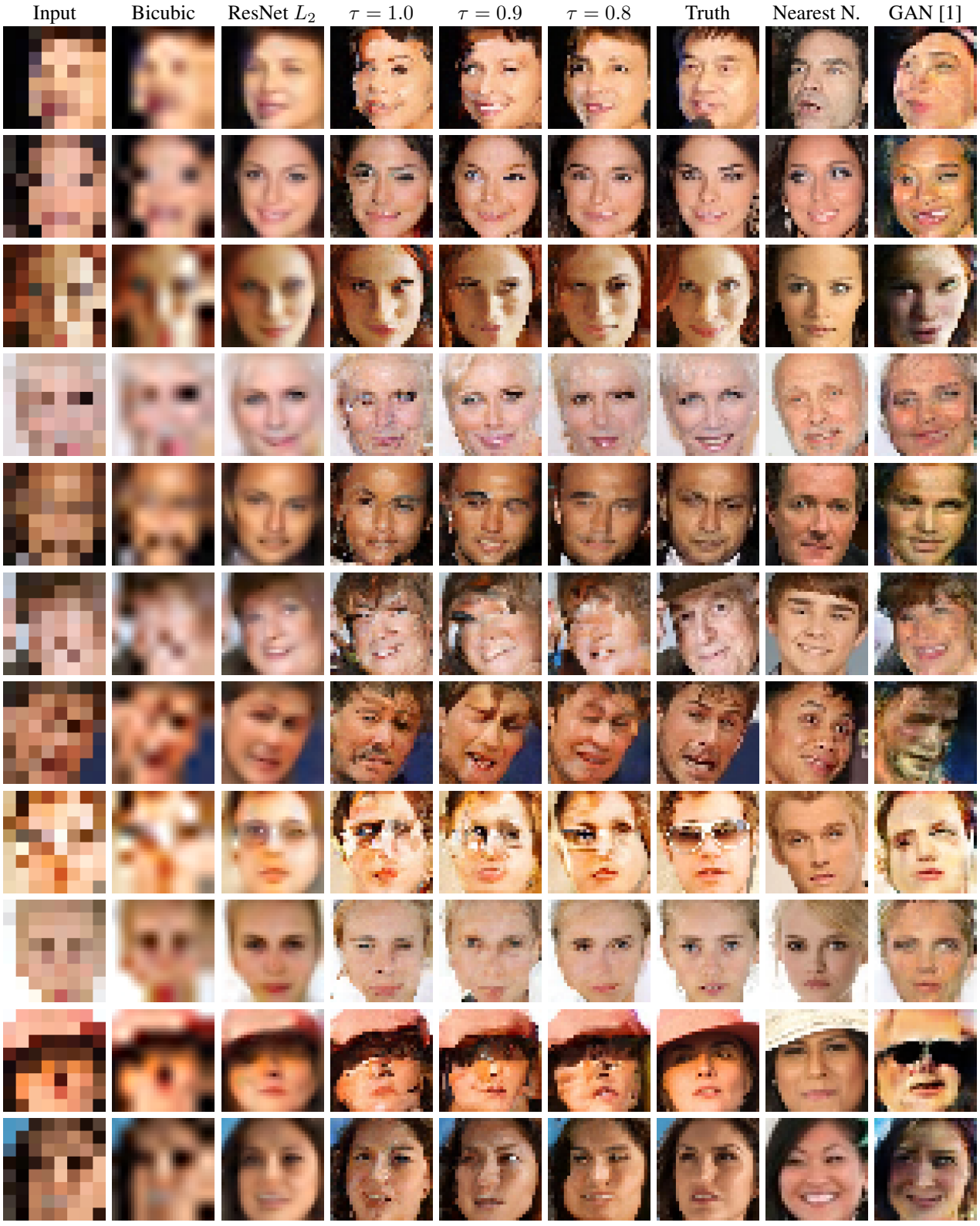

Input
Bicubic
ResNet L_2
$\tau = 1.0$
$\tau = 0.9$
$\tau = 0.8$
Truth
Nearest N.

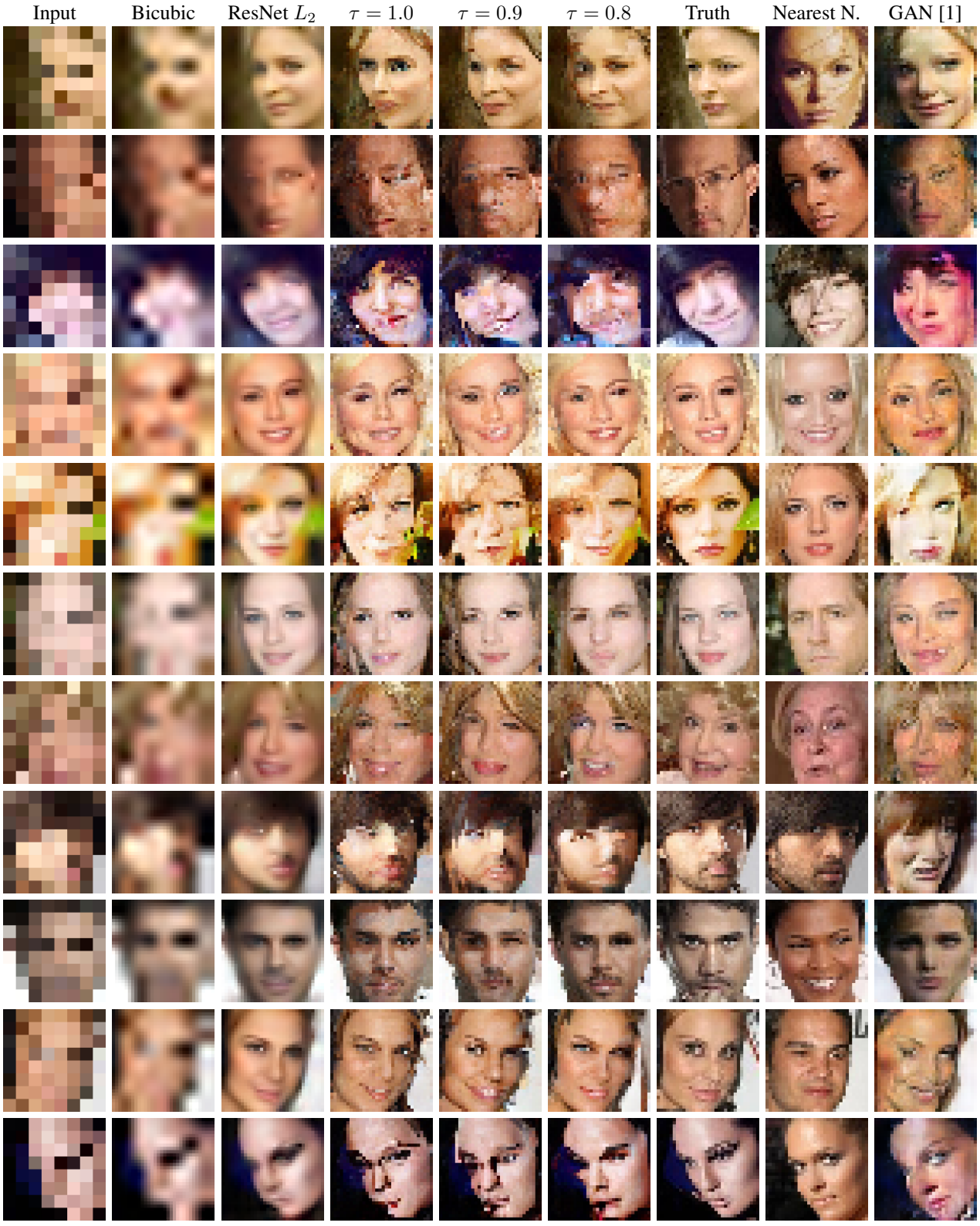
3. Samples from models trained on CelebA faces




Input
Bicubic
ResNet L_2
$\tau = 1.0$
$\tau = 0.9$
$\tau = 0.8$
Truth
Nearest N.
GAN [1]


Input
Bicubic
ResNet L_2
$\tau = 1.0$
$\tau = 0.9$
$\tau = 0.8$
Truth
Nearest N.
GAN [1]


Input
Bicubic
ResNet L_2
$\tau = 1.0$
$\tau = 0.9$
$\tau = 0.8$
Truth
Nearest N.
GAN [1]


Input
Bicubic
ResNet L_2
$\tau = 1.0$
$\tau = 0.9$
$\tau = 0.8$
Truth
Nearest N.
GAN [1]

4. Samples images that performed best and worst in human ratings.

The best and worst rated images in the human study. The fractions below the images denote how many times a person choose that image over the ground truth.

References

[1] D. Garcia. srez: Adversarial super resolution. 2016.