|  | Patch 0 | Reference | Patch 1 | Patch 0 | Reference | Patch 1 | Patch 0 | Reference | Patch 1 |
|---|---|---|---|---|---|---|---|---|---|
| Humans |  |  | ✔ | ✔ |  |  | ✔ |  |  |
| L2/PSNR, SSIM, FSIM | ✔ |  |  |  |  | ✔ |  |  | ✔ |
| Random Networks | ✔ |  |  |  |  | ✔ |  |  | ✔ |
| Unsupervised Networks |  |  | ✔ | ✔ |  |  | ✔ |  |  |
| Self-Supervised Networks |  |  | ✔ | ✔ |  |  | ✔ |  |  |
| Supervised Networks |  |  | ✔ | ✔ |  |  | ✔ |  |  |

Figure 1: **Which patch (left or right) is "closer" to the middle patch in these examples?** In each case, the traditional metrics (L2/PSNR, SSIM, FSIM) disagree with human judgments. But deep networks, even across architectures (Squeezenet [20], AlexNet [27], VGG [52]) and supervision type (supervised [47], self-supervised [13, 40, 43, 64], and even unsupervised [26]), provide an *emergent embedding* which agrees surprisingly well with humans. We further cal-