Google

# Revisiting spatial invariance with low rank local connectivity

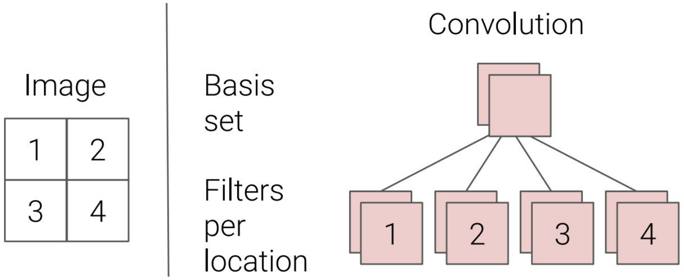
Gamaleldin Elsayed, Prajit Ramachandran, Jonathon Shlens, Simon Kornblith

Google Research, Brain Team




# Is spatial invariance a good inductive bias?

- Convolutional architectures perform better than locally connected on computer vision problems.
- Both convolution and local connectivity assume local receptive fields as an inductive bias.
- Distinction between the two is requiring **spatial invariance** in convolution.
- **Spatial invariance:** local filter bank is shared and applied equally across space.



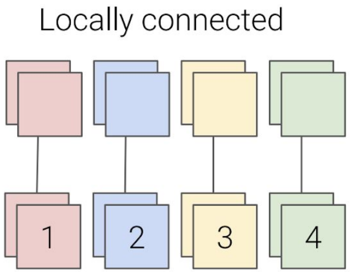




Image from https://opidesign.net/landscape-architecture/landscape-architecture-fun-facts/

# Is spatial invariance a good inductive bias?

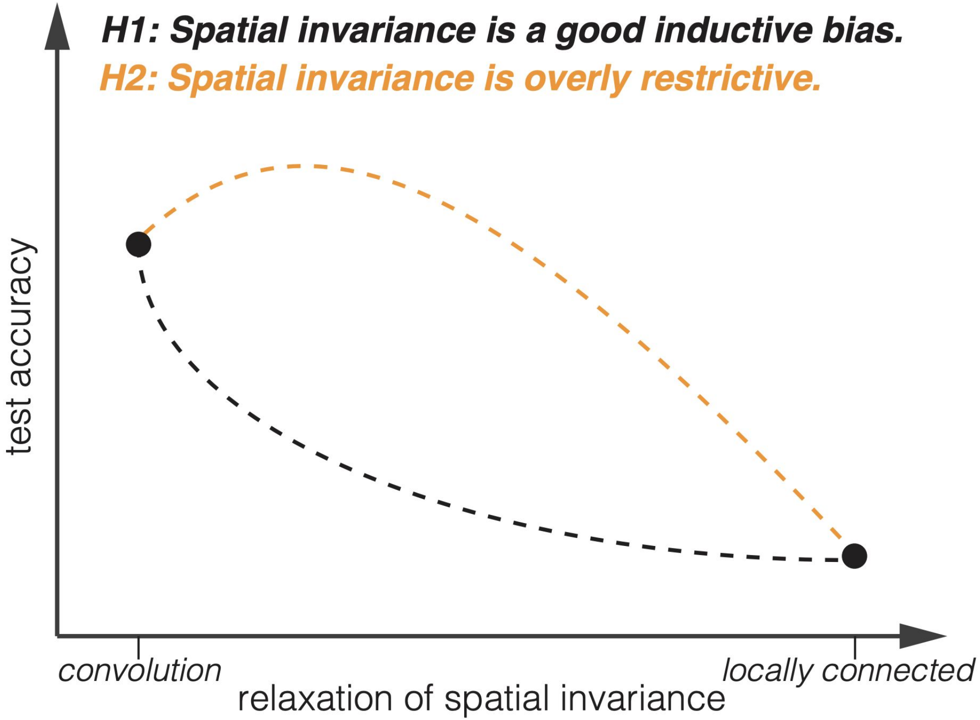
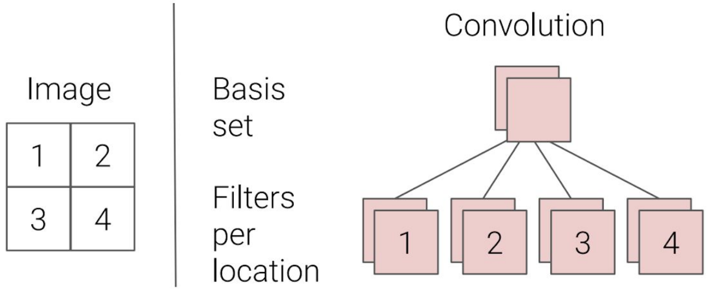


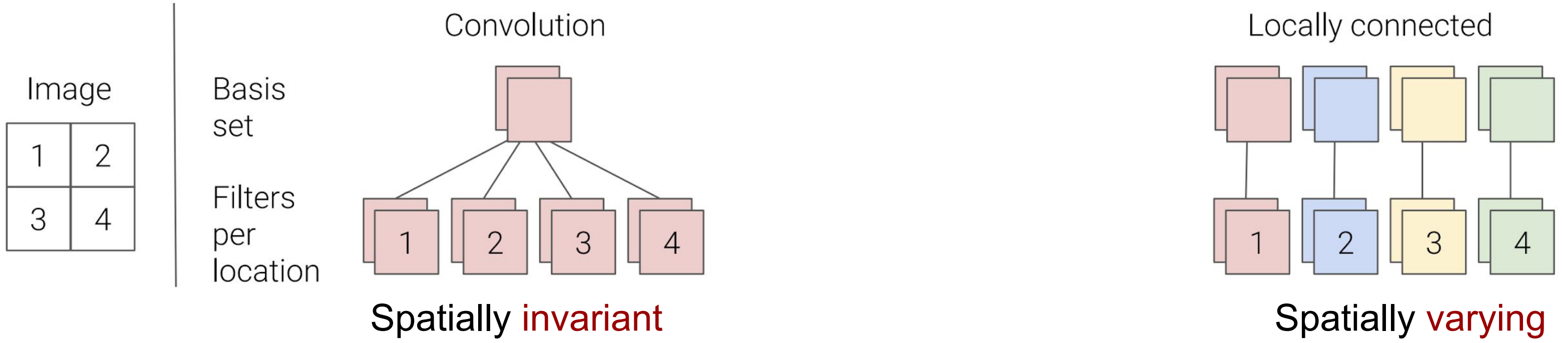

Image from
https://opidesign.net/landscape-architecture/landscape-architecture-fun-facts/

# Low rank local connectivity (LRLC)

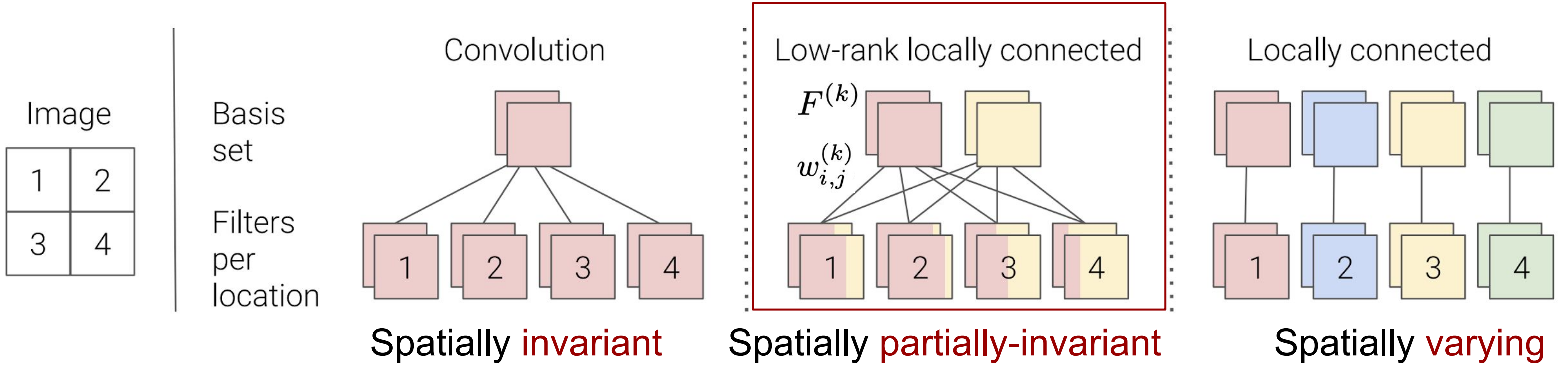



Spatially invariant

# Low rank local connectivity (LRLC)



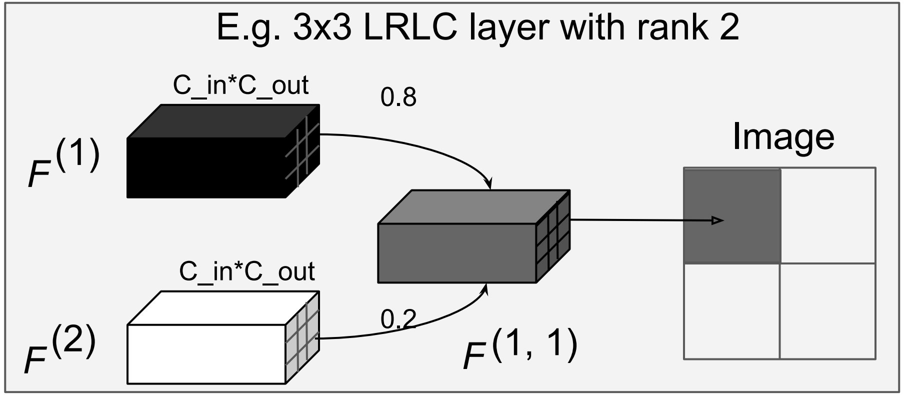
# Low rank local connectivity (LRLC)



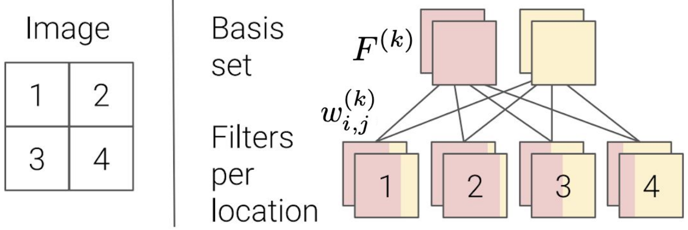

- Basis set of K local filter banks (controls the degree of relaxation of spatial invariance):

$$\{F^{(1)}, \ldots, F^{(K)}\} \in K\mathbb{R}^{h \times w \times C_{\text{in}} \times C_{\text{out}}} \text{ (basis set)}$$

$$F^{(i,j)} = \sum_{k=1}^{K} w_{i,j}^{(k)} F^{(k)}$$



# Types of LRLC layers




$$w_{i,j}^{(k)} = \frac{\exp\left(\tilde{w}_{i,j}^{(k)}\right)}{\sum_{l=1}^{K} \exp\left(\tilde{w}_{i,j}^{(l)}\right)}$$

| Fixed LRLC | Input-dependent LRLC |
|---|---|
| Fixed basis set of K filter banks. $\{F^{(1)}, \ldots, F^{(K)}\} \in K\mathbb{R}^{h \times w \times C_{\text{in}} \times C_{\text{out}}}$ | Fixed basis set of K filter banks. $\{F^{(1)}, \ldots, F^{(K)}\} \in K\mathbb{R}^{h \times w \times C_{\text{in}} \times C_{\text{out}}}$ |
| Fixed combining weights $\tilde{w}_{i,j}^{(k)}$. | Combining weights are generated by a simple neural network $\tilde{w}_{i,j}^{(k)} = g_{i,j}^{(k)}(I)$. |
| Learnable parameters:<br>K filter banks and combining weights. | Learnable parameters:<br>K filter banks and the simple network parameters. |

# Experiments

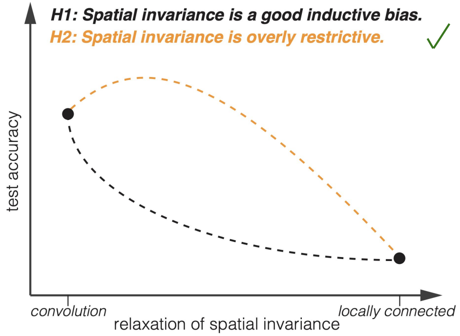


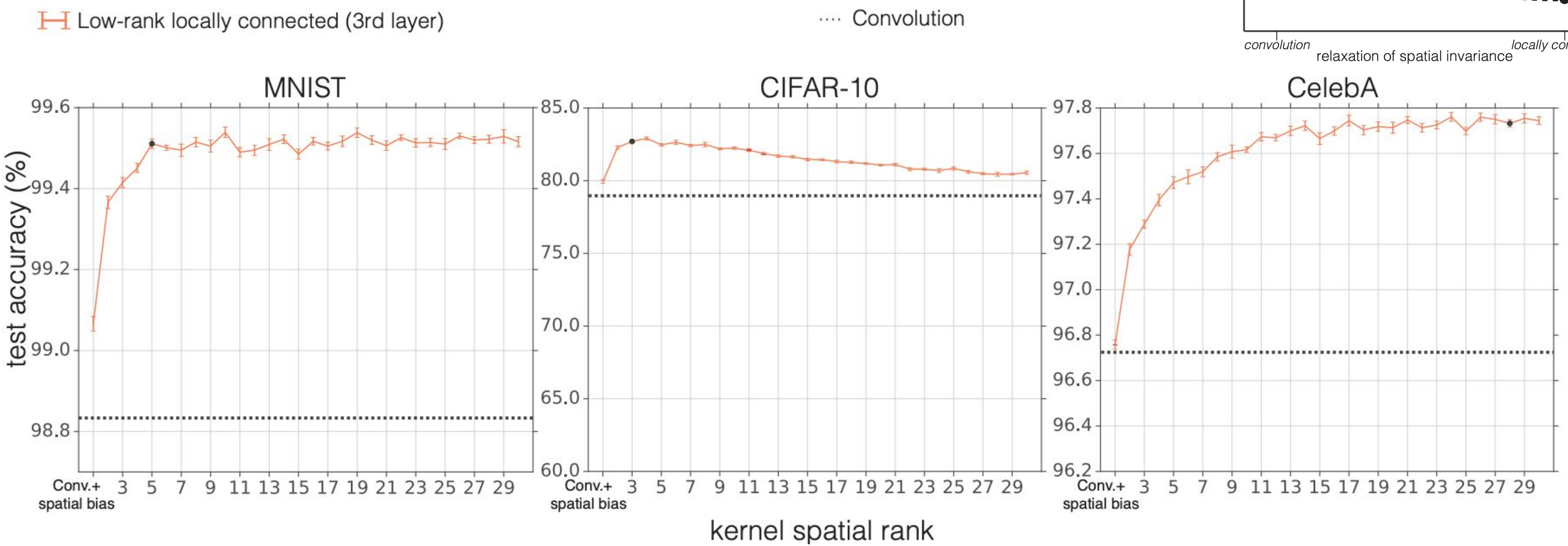
- Datasets:
  - MNIST.
  - CIFAR-10.
  - CelebA.

- Network: 3 layer network with 3x3 filter sizes and 64 channels (global average pooling with fully connected).
- No augmentation or regularization to focus on architecture effects.
- We also demonstrate the feasibility of applying LRLC to large scale problems by running experiments on ImageNet.

# Spatial invariance may be overly restrictive

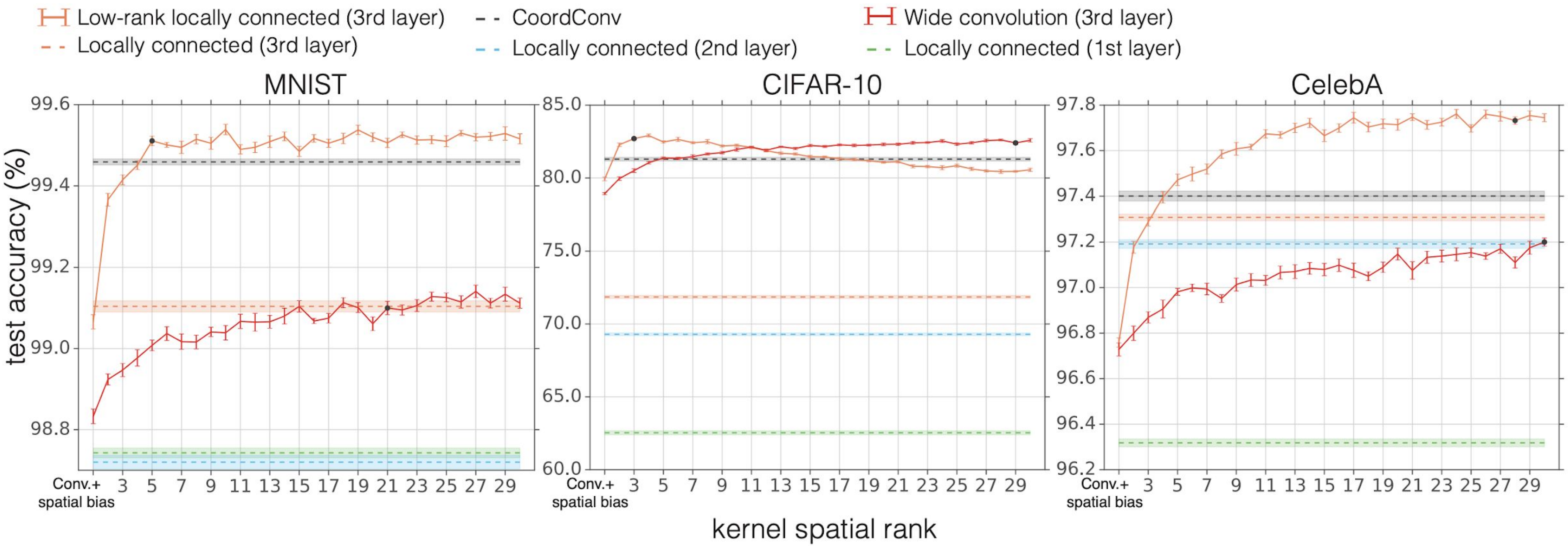
Accuracy increases over convolution baseline as we relax spatial invariance consistent with our hypothesis.



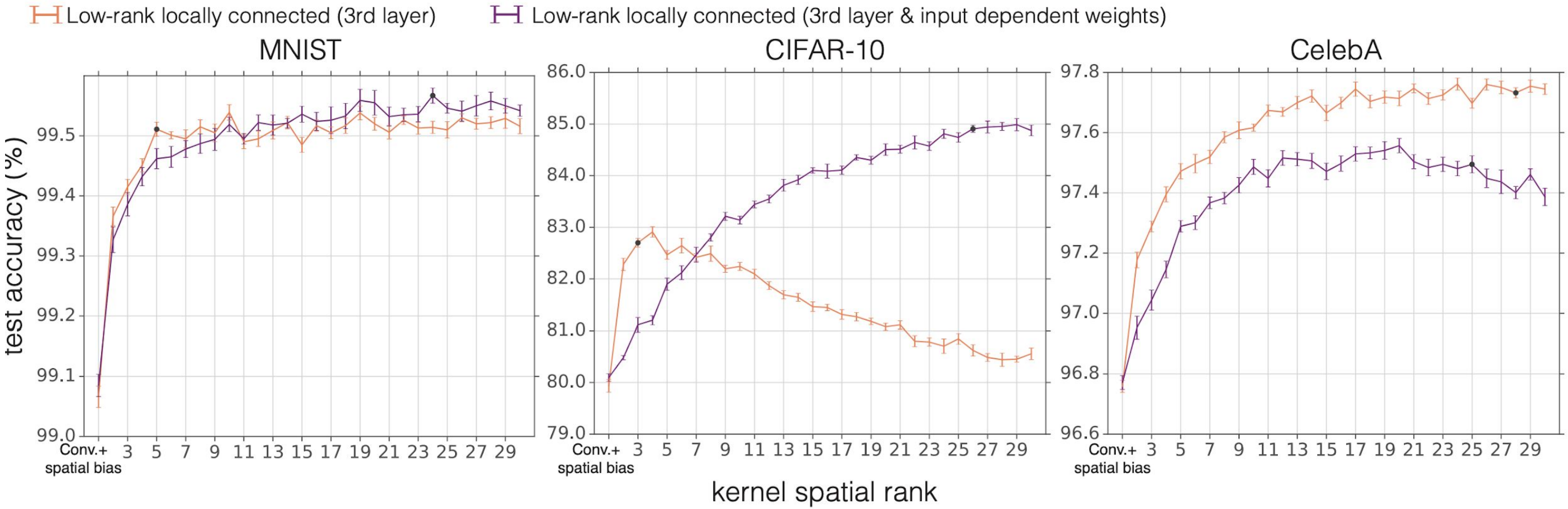
# Spatial invariance may be overly restrictive

Low rank local connectivity outperforms wide convolutions, locally connected layers, and coord conv.
Optimal rank is dataset dependent and is higher for more aligned data (eg CelebA) than less aligned data (CIFAR-10).



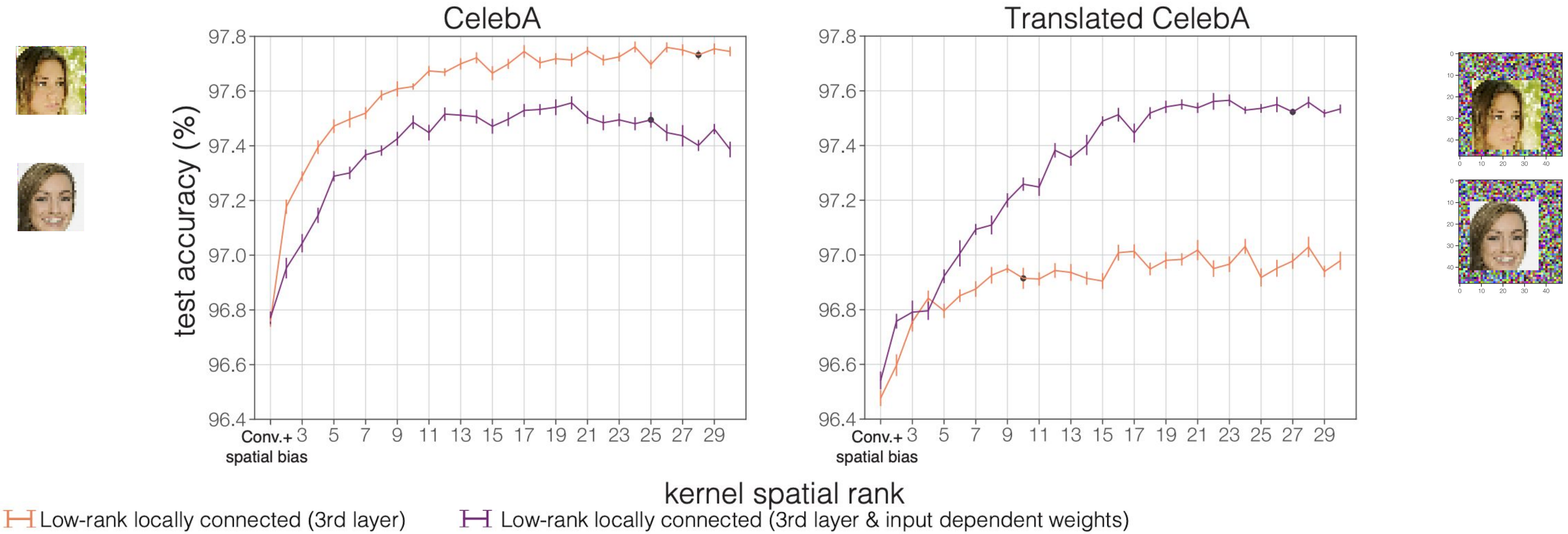
# Input-dependent LRLC is a better inductive bias for datasets with less alignment

Less aligned dataset: Input-dependent LRLC suits CIFAR-10 better than fixed LRLC.
More aligned dataset: Fixed LRLC suits CelebA better than input-dependent LRLC.

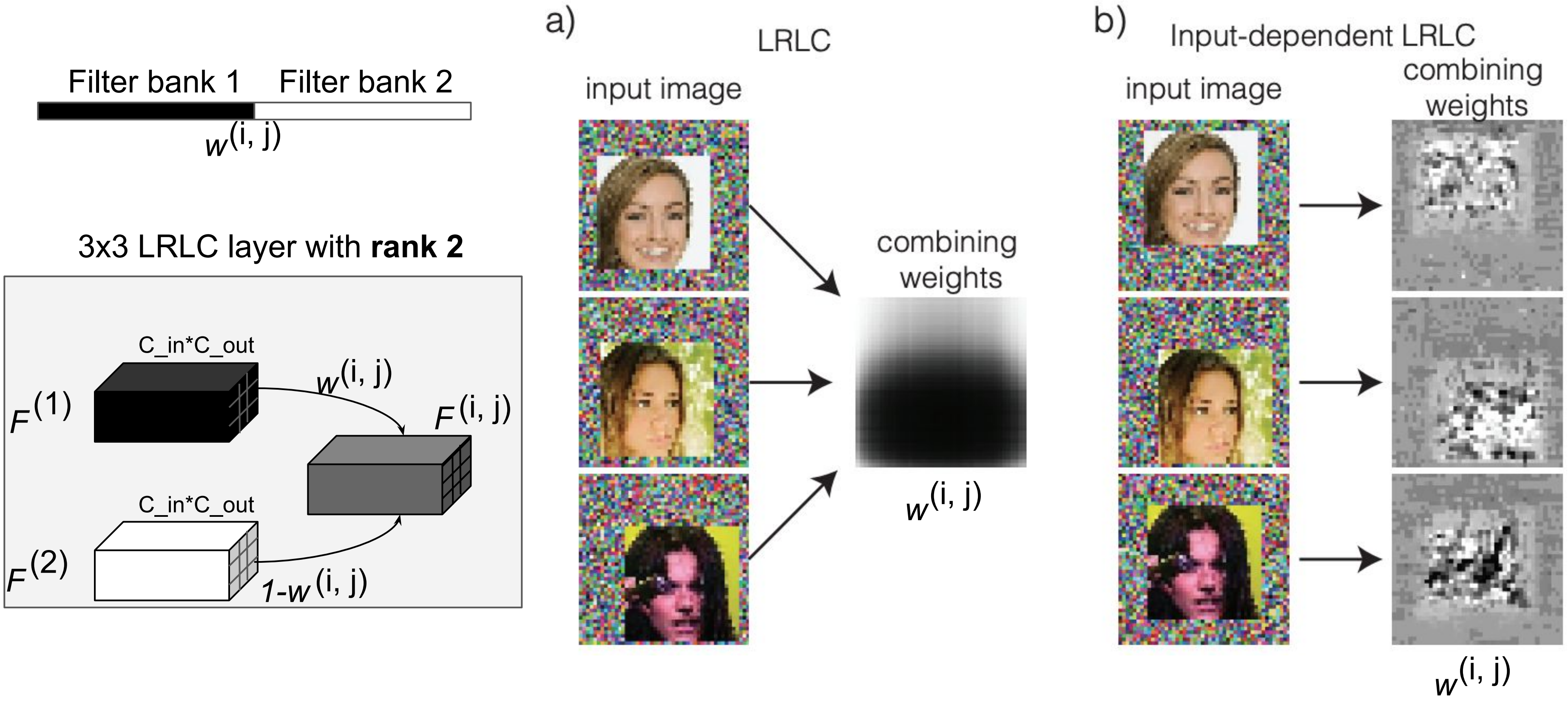


# Input-dependent LRLC is a better inductive bias for datasets with less alignment

Misaligned examples in translated CelebA impact the fixed LRLC model performance but not the input-dependent LRLC.



# Visualization of learned combining weights



# Feasibility of the application of LRLC to large scale problems

- Locally connected layers are prohibitively expensive to apply to large scale problems.
- Parameter count of the LRLC layer scales only with rank, making it feasible to apply to large scale problems.
- We demonstrate this feasibility by applying LRLC to ResNet-50 on ImageNet 224x224.

| LAYER | INSERT LAYER | REPLACE 3X3 CONVS |
|---|---|---|
| CONVOLUTION | $77.22 \pm 0.03$ | $76.93 \pm 0.07$ |
| COORDCONV(LIU ET AL., 2018) | $77.23 \pm 0.03$ | $77.07 \pm 0.08$ |
| LRLC | $77.47 \pm 0.03$ | $77.08 \pm 0.02$ |
| LRLC (INPUT DEPENDENT WEIGHTS) | $77.45 \pm 0.03$ | $77.80 \pm 0.02$ |
| WIDE CONVOLUTION | **77.48±0.05** | **78.54±0.04** |

# Conclusions

- We design a new layer (LRLC) that can parametrically adjust the degree of spatial invariance to test whether spatial invariance is a good inductive bias.
- **Main takeaway:** we demonstrate that spatial invariance in convolutional layers may be an overly restrictive inductive bias.
- Unlike locally connected layers, parameter count of the LRLC layer scales only with rank, making it feasible to apply to large scale problems.
- Future direction: applications of LRLC to other computer vision problems.

# Acknowledgements

We thank the following for useful discussions and helpful feedback on the paper:
Jiquan Ngiam
Pieter-Jan Kindermans
Jascha Sohl-Dickstein
Jaehoon Lee
Daniel Park
Sobhan Naderi
Max Vladymyrov
Hieu Pham
Michael Simbirsky
Roman Novak
Hanie Sedghi
Karthik Murthy
Michael Mozer
Yani Ioannou

# Questions?

Thank you!

Paper: https://arxiv.org/abs/2002.02959

Code: https://github.com/google-research/google-research/tree/master/low_rank_local_connectivity