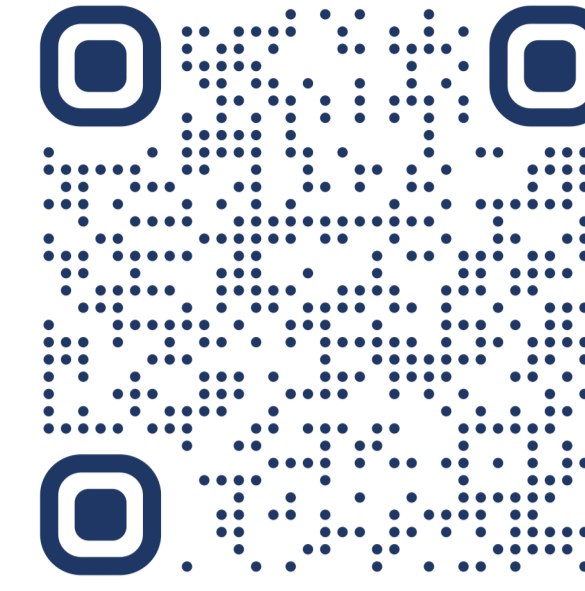


INAM: Image-Scale Neural Additive Models

Jana Hüls, Jan-Ole Perschewski, Sebastian Stober

Otto-von-Guericke-University – Artificial Intelligence Lab

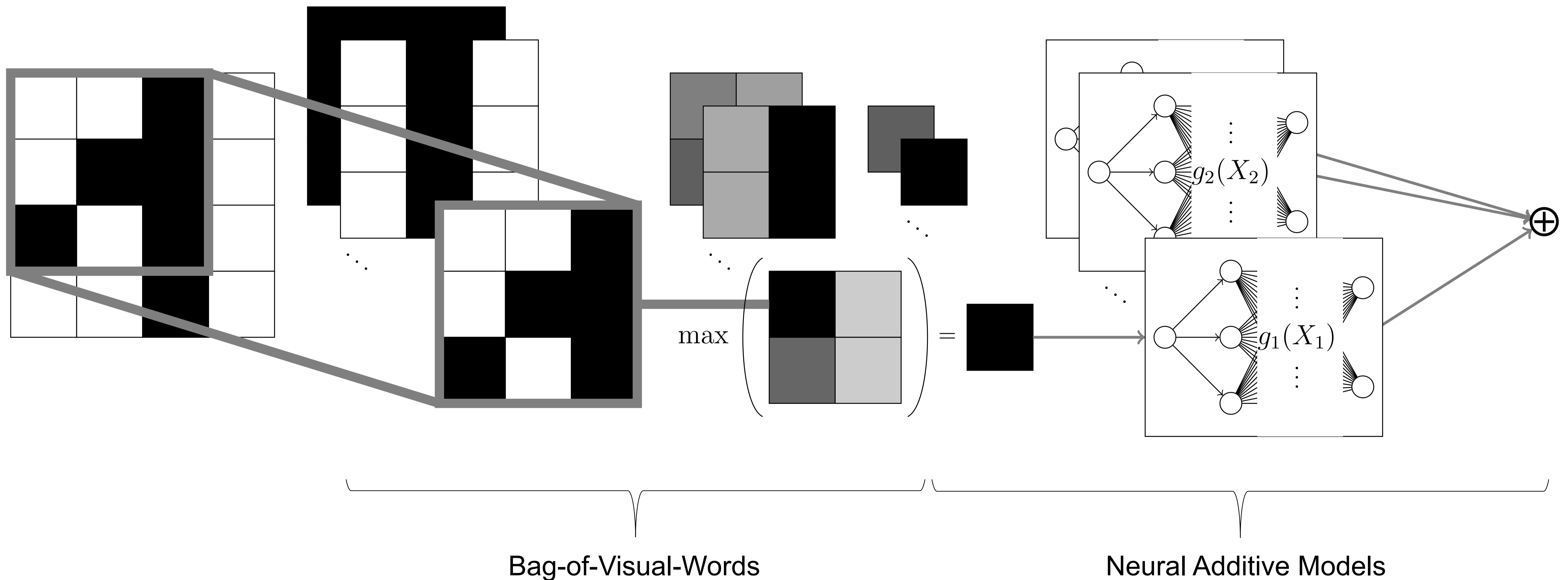
contact: {jan-ole.perschewski, stober}@ovgu.de



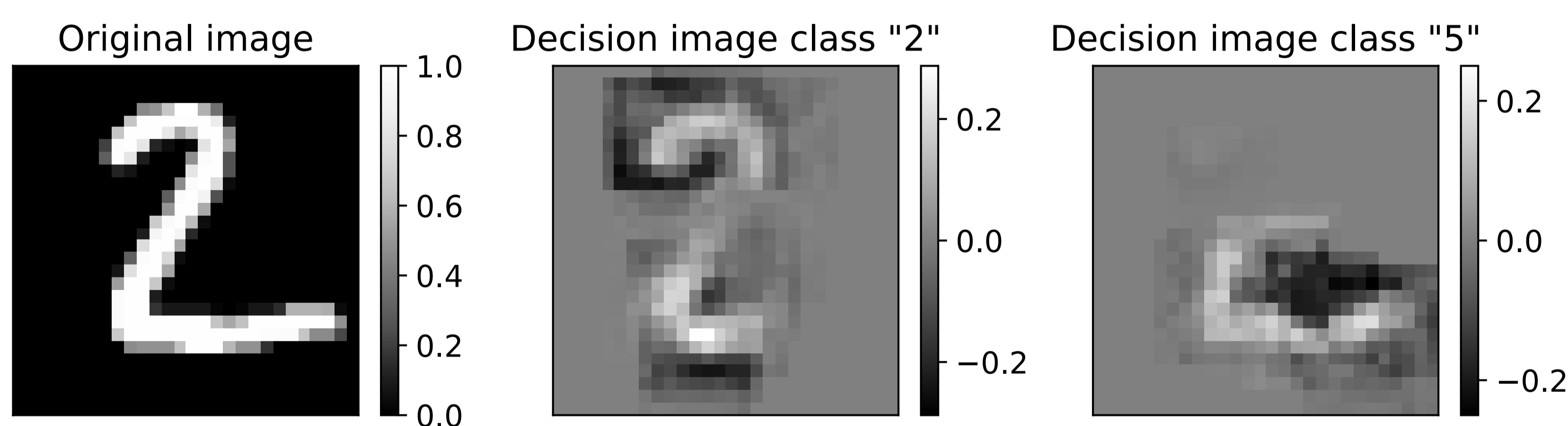
Motivation

Deep neural networks are not interpretable → Make neural networks interpretable by design

INAM

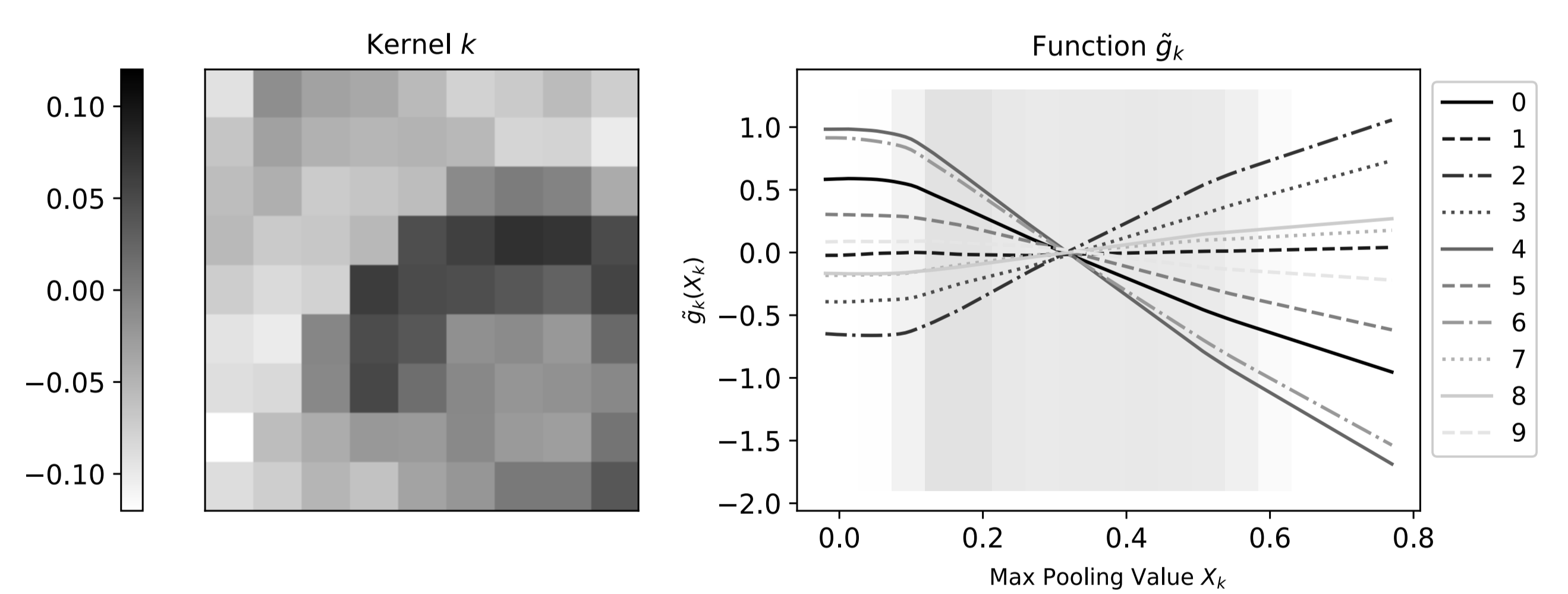


Local Explanations



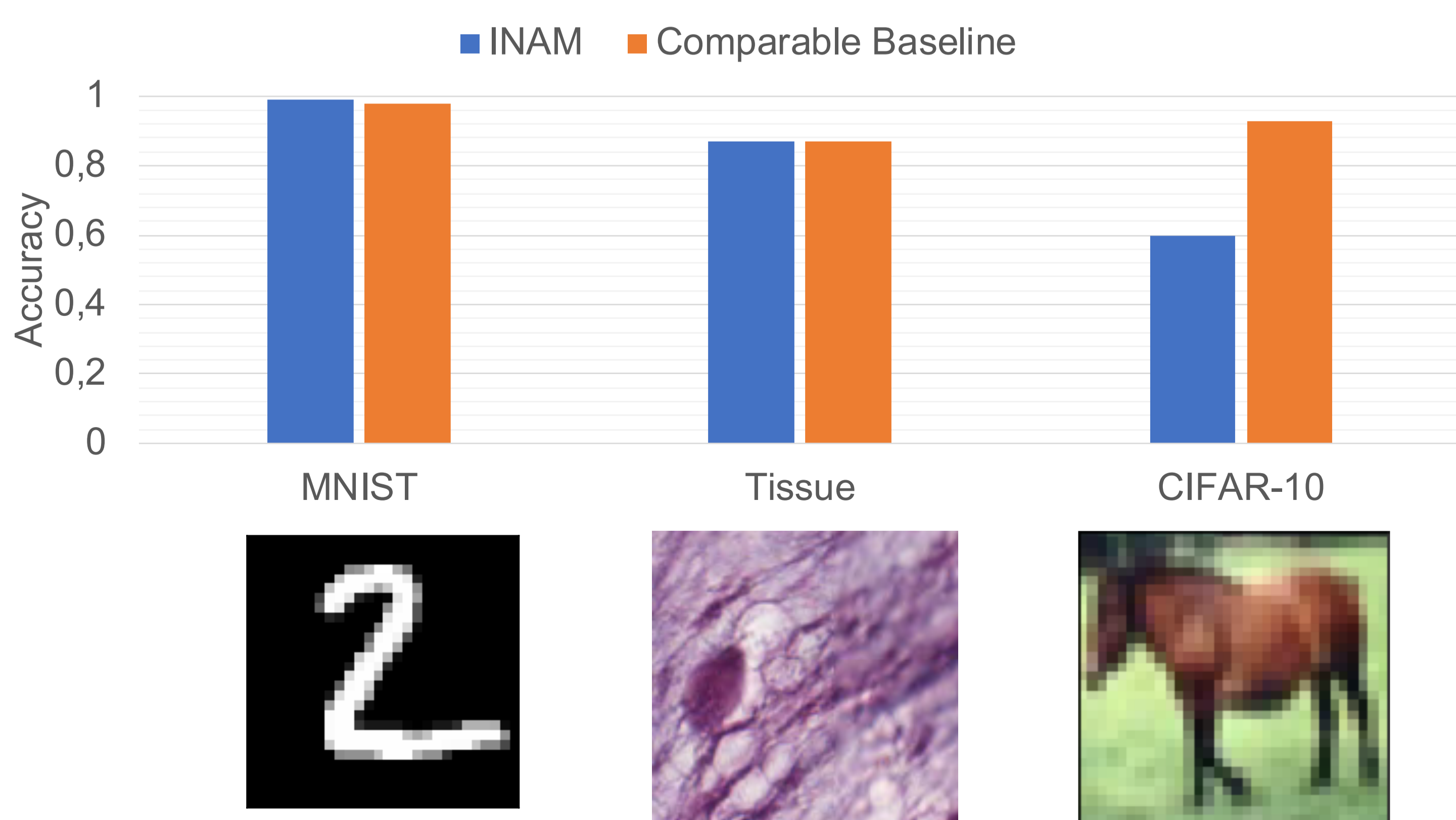
Sum over kernels at maximum position weighted by their contribution for a given class

Feature Influence

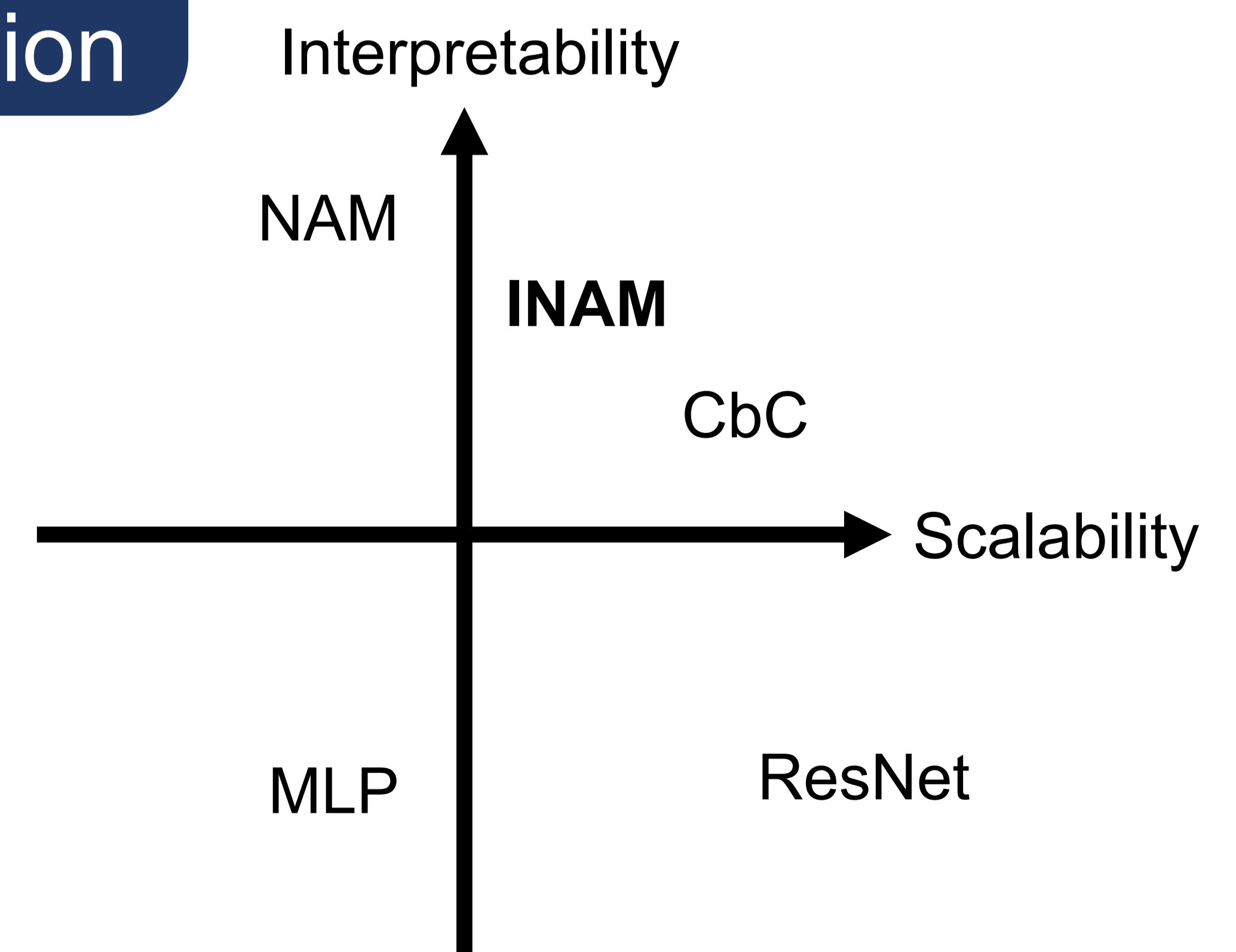


Neural Additive Model Visualization with respect to the detected feature

Performance



Conclusion



How can this approach be scaled to complex data sets without adding too many parameters?