Towards model-driven neural networks
Lecture Series Artificial Intelligence

Department of Numerical Analysis, Faculty of Informatics,

ELTE Eötvös Loránd University, Budapest, Hungary

Péter Kovács

December 15, 2020

Outline

- Motivations
- Thermographic imaging
- Model based approach
- ML based approach
- Hybrid approach
- Experiments
- VPNet
- Architectures
- Experiments
- Conclusion

- Motivations
- Thermographic imaging
- Model based approach
- ML based approach
- Hybrid approach
- Experiments
- VPNet
- Architectures
- Experiments
- Conclusion

Past

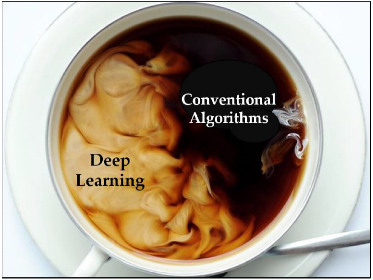
Model-based algorithms originated by humans dominate

Present

Data-driven blackbox DNNs take the lead in many fields

Future

Model-driven DNNs possibly dominate in signal processing

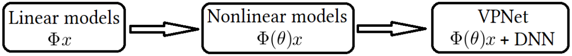


Why to combine model-based algorithms with DNN?

- Improve the performance: incorporate domain knowledge.
- Open the blackbox: interpret DNN predictions.
- Temper DNN's data-hunger: reduce the training data.

How to do that?

- Traditional approach: feature engineering.
  - Pros: dimensionality reduction, interpretability.
  - Cons: usually suboptimal.
- Architectures based on existing algorithms: deep unfolding, Wiener–Hammerstein type NNs, spline adaptive filtering, tensor-based learning.
- Representation learning: convolutional nerual networks, autoencoders, etc.



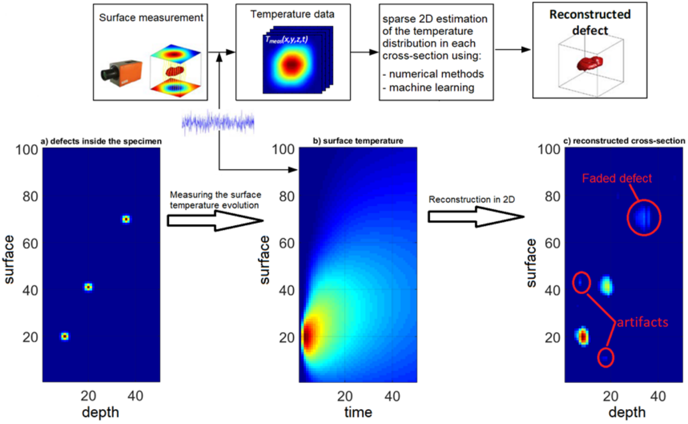
Case studies

- Thermographic image regression
  - Linear models + DNN
  - Domain knowledge included via physical models
- ECG classification
  - Nonlinear models + DNN
  - Domain knowledge included via model-driven representation learning

Problem description
Thermographic imaging

Goal

Analysis, detection of structural imperfections of materials.



Linear model

- $\mathbf{d}$: noisy surface temperature measurements after heating.
- $\mathbf{u}$: initial temperature distribution inside the material.
- $\mathbf{\Phi}$: forward mapping that models the heat conduction.
- The corresponding discrete linear inverse problem:
$$\mathbf{\Phi u} = \mathbf{d}.$$

Heuristics

- Nonnegativity: entries of $\mathbf{u}$ represents temperature data.
- Sparsity: locations of nonzeros in $\mathbf{u}$ are either defects or noise.
- Group sparsity: nonzero groups in $\mathbf{u}$ are most probably defects.

Challenges in thermographic imaging

- Numerical: it is a discrete ill-posed inverse problem.
- Computational: it is a large-scale problem.
- Modeling: how to derive $\mathbf{\Phi}$?

Two-stage reconstruction process [1]

1) Transformation of the thermographic imaging problem:

$$\widetilde{\mathbf{v}} = \arg\min_{\mathbf{v}}\{\|\mathbf{d} - \mathbf{K}\mathbf{v}\|_2^2 + \lambda^2 \cdot \Omega(\mathbf{v})\}.$$

2) Applying ultrasonic imaging techniques to the new problem:

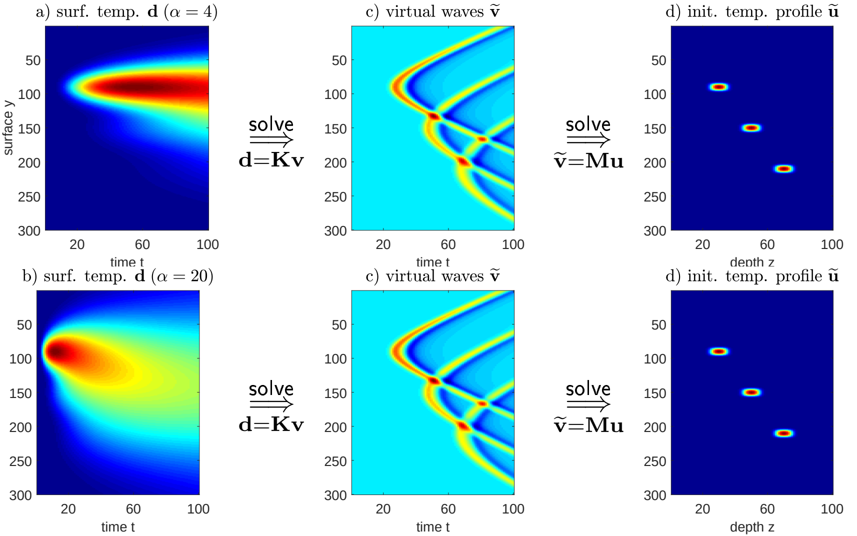
$$\widetilde{\mathbf{u}} = \arg\min_{\mathbf{u}}\{\|\widetilde{\mathbf{v}} - \mathbf{M}\mathbf{u}\|_2^2 + \mu^2 \cdot \Omega(\mathbf{u})\}.$$

One-stage reconstruction process

By $\mathbf{\Phi} = \mathbf{K}\mathbf{M}$, the full reconstruction can be written as follows:

$$\widetilde{\mathbf{u}} = \arg\min_{\mathbf{u}}\{\|\mathbf{d} - \mathbf{\Phi}\mathbf{u}\|_2^2 + \nu^2 \cdot \Omega(\mathbf{u})\}.$$

[1] P. Burgholzer, M. Thor, J. Gruber, and G. Mayr. Three-dimensional thermographic imaging using a virtual wave concept. Journal of Applied Physics, 121(10):105102 1–11, 2017.



a) surf. temp. $\mathbf{d}$ ($\alpha = 4$)
b) surf. temp. $\mathbf{d}$ ($\alpha = 20$)
c) virtual waves $\tilde{\mathbf{v}}$
c) virtual waves $\tilde{\mathbf{v}}$
d) init. temp. profile $\tilde{\mathbf{u}}$
d) init. temp. profile $\tilde{\mathbf{u}}$
surface y
time t
depth z
solve $\Longrightarrow$ $\mathbf{d}$=$\mathbf{Kv}$
solve $\Longrightarrow$ $\tilde{\mathbf{v}}$=$\mathbf{Mu}$
50
100
150
200
250
300
20
60
100

Two-stage reconstruction process

Model based approach

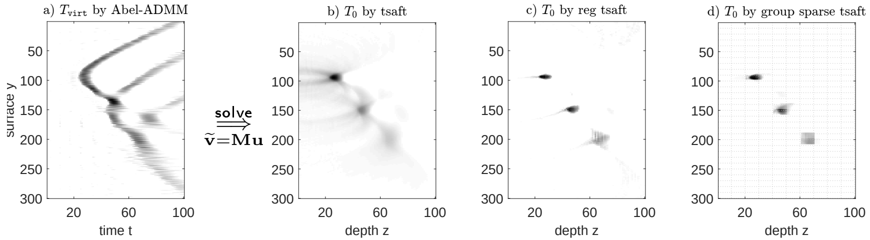


Figure: Second stage of the reconstruction process. *a)* Virtual wave reconstruction by ADMM with Abel trf.; initial temperature distribution by *b) tsaft*, *c) reg tsaft*, and *d)* group sparse *grp. tsaft*, where groups of size 10×10 were used as indicated by the black grid.

Motivation

- Existing iterative algorithms can converge to bad solutions.
- The convergence can be slow, and the solution have to be recalculated for every new image.
- Incorporating problem-specific information into an algorithm.

Sparse estimation via ML techniques

We want to learn an algorithm for solving

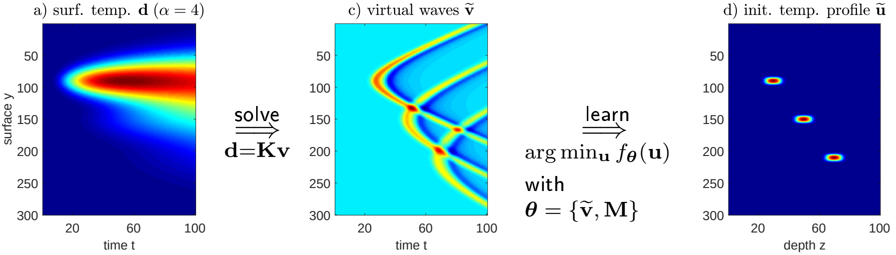
$$\widetilde{\mathbf{u}} = \arg\min_{\mathbf{u}} f_{\boldsymbol{\theta}}(\mathbf{u}) = \arg\min_{\mathbf{u}} \{\|\mathbf{d} - \boldsymbol{\Phi}\mathbf{u}\|_2^2 + \tau^2 \cdot \|\mathbf{u}\|_1\}.$$

Parameter class of interests

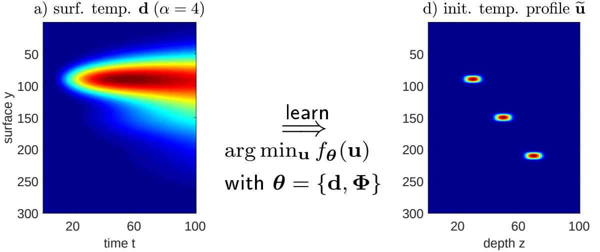
$$\boldsymbol{\theta} = \{\mathbf{d}, \boldsymbol{\Phi}\}, \quad \Omega = \left\{\mathbf{d}, \boldsymbol{\Phi} \,|\, \mathbf{d} \in \mathbb{R}^N, \, \boldsymbol{\Phi} \in \mathbb{R}^{N \times M}\right\}$$

$\boldsymbol{\Phi}$ and $\mathbf{d}$ are not arbitrary, they are prescribed by physical models.

Two-stage reconstruction by the hybrid ML approach.



One-stage reconstruction by the end-to-end ML approach.



Reconstruction in 2D

1) Extract the virtual waves $\widetilde{\mathbf{v}}$ from the measurements $\mathbf{d}$.
   - utilize the sparse and non-negative nature of $\widetilde{\mathbf{v}}$;
2) Estimate the temperature distribution $\widetilde{\mathbf{u}}$ by machine learning:
   - input: thermal diffusivity invariant virtual waves $\widetilde{\mathbf{v}}$
   - output: approximation of $\widetilde{\mathbf{u}}$

Reconstruction in 3D [2]

- Estimate the temperature distribution in each 2D cross-section.
- 3D reconstruction from the sequence of 2D images.

[2] P. Kovács, B. Lehner, G. Thummerer, G. Mayr, P. Burgholzer, M. Huemer, Deep learning approaches for thermographic imaging, Journal of Applied Physics, 2020, vol. 128, no. 15, pp. 155103-1-16.

Deep learning by u-net

Hybrid approach

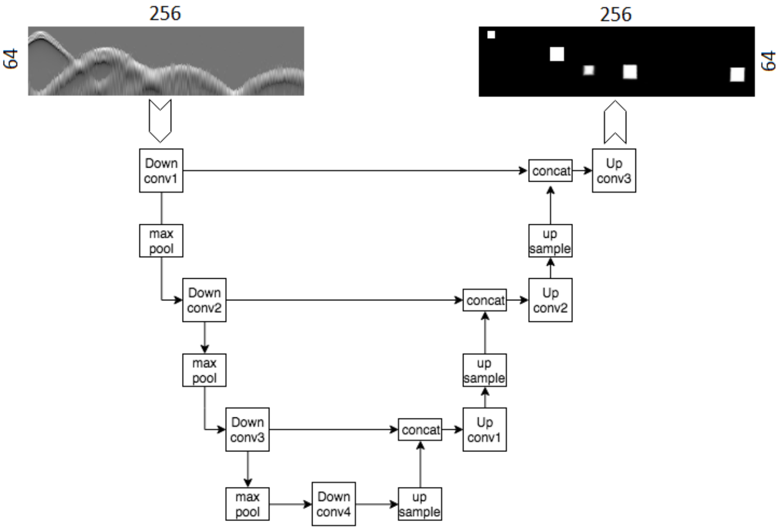


Figure: Architecture of the compact u-net.

Mathematical model of the heat diffusion

$$\left(\nabla^2 - \frac{1}{\alpha}\frac{\partial}{\partial t}\right) T(\mathbf{r}, t) = -\frac{1}{\alpha} T_0(\mathbf{r})\delta(t),$$

where

- α stands for the thermal diffusivity,
- T is the temperature as a function of space $\mathbf{r}$ and time t,
- T_0 denotes the initial temperature profile at $t = 0$.

Data generation in 2D assuming adiabatic boundary conditions

$$\widehat{T}(k_y, k_z, t) = \widehat{T}_0(k_y, k_z) \cdot \exp(-(k_y^2 + k_z^2) \cdot \alpha t),$$

where

- $\widehat{T}$, $\widehat{T}_0$ are the cosine transforms of T and T_0 in the yz-plane,
- k_y and k_z are the corresponding spatial frequencies.

Data sets
Hybrid approach

Training data

- 8,000 simulated noise free samples with adiabatic boundary conditions.
- 2-5 square-shaped defects with side lengths between 2 and 6 pixels.
- The resolution of each image is 256×64.
- 10 different versions of each sample were used, representing SNRs from -20 dB to 70 dB in 10 dB steps.
- Overall number of training images: 10×8000

Testing data

- 1,000 simulated samples similar to the training images.
- Overall number of test images: 10×1000
- Real measurement data containing 256 images of size 256×64.

- Motivations
- Thermographic imaging
- Model based approach
- ML based approach
- Hybrid approach
- Experiments
- VPNet
- Architectures
- Experiments
- Conclusion

Numerical solvers for sparse approximation

- SPGL1 is for large-scale one-norm regularized least squares.
- YALL1 is a solver for basic/group sparse reconstruction.
- ASP is for solving several variations of the sparse optimization.
- **ADMM** (alternating direction method of multipliers) is a very general algorithm for solving sparse approximation problems.
- SALSA is a fast ADMM type algorithm for image reconstruction.
- **IRfista** is a recent numerical solver for large-scale problems.

Tested model based approaches

- fkmig: Stolt's f-k migration without sparse regularization.
- tsaft: Synthetic Aperture Focusing Technique in the time domain.
- reg tsaft: same as tsaft, but with sparse regularization.

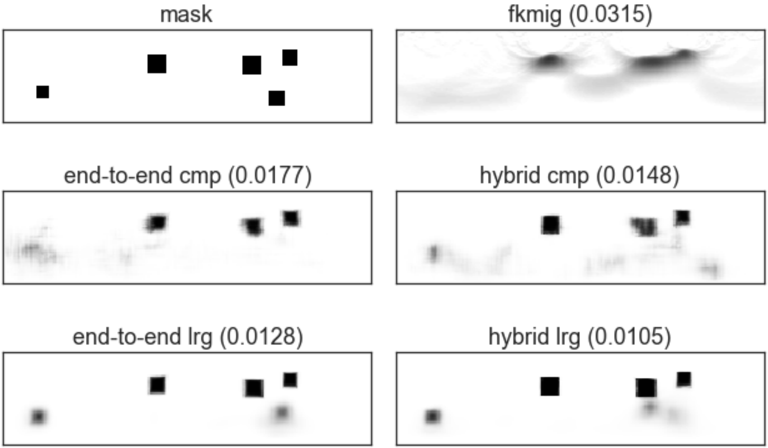


Figure: Reconstructions of a 0 dB SNR example from the test set.

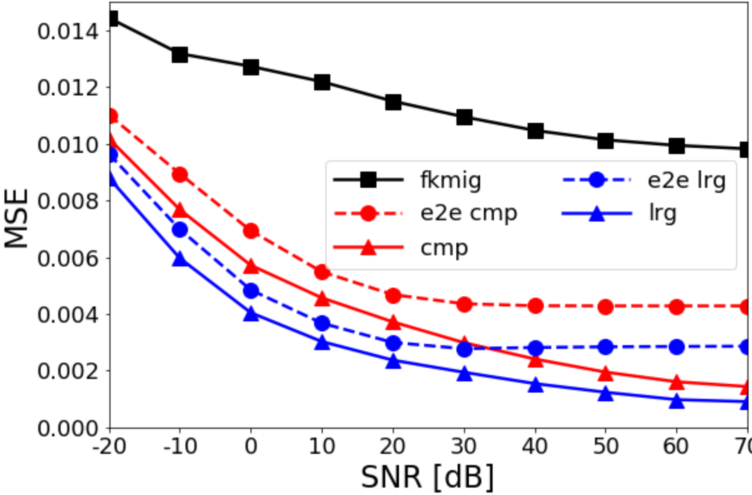


Figure: The MSE of the baselines and the proposed method.

Real measurement data

Experiments

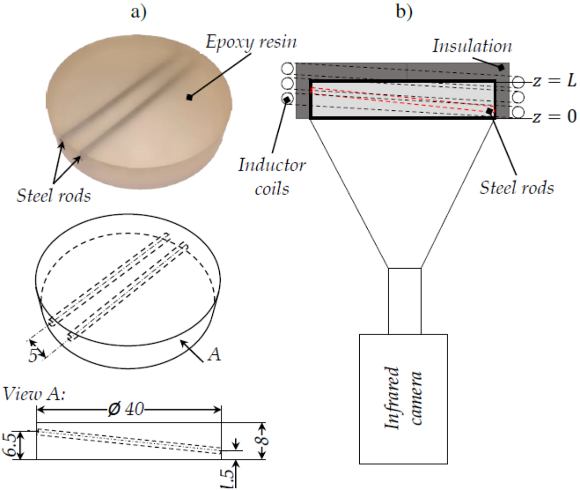


Figure: Parameters of the phantom.

Real measurement data

Experiments

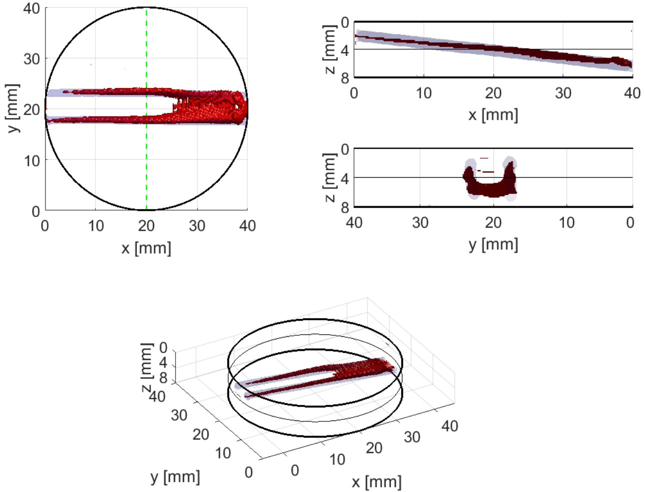


Figure: Using the model-based *fkmig* approach for 3D reconstruction of the specimen without rotation.

Real measurement data

Experiments

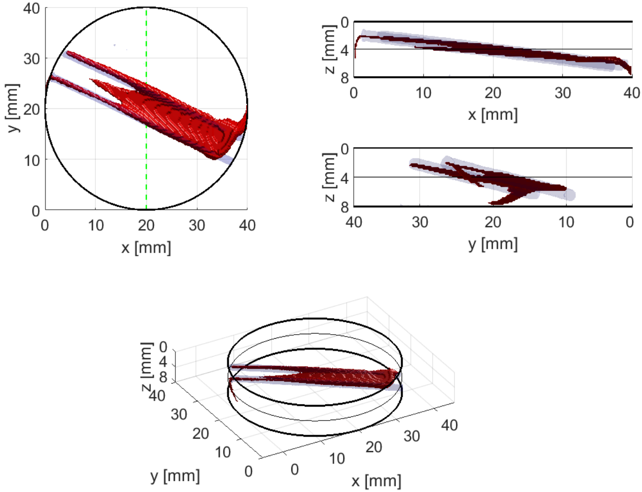


Figure: Using the model-based *fkmig* approach for 3D reconstruction of the specimen a rotation of $25°$.

Real measurement data

Experiments

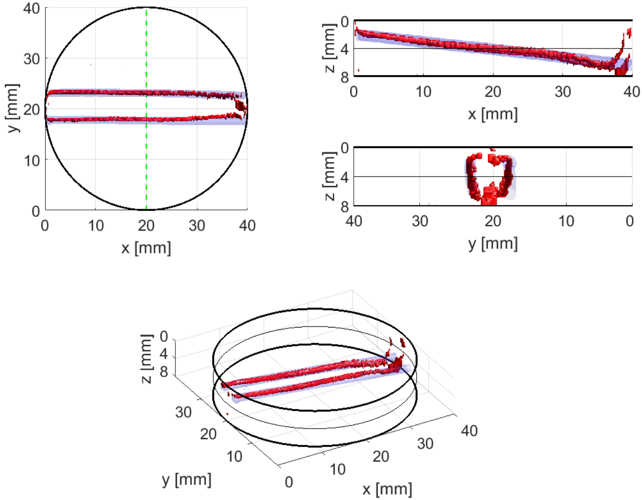


Figure: Using the large end-to-end *e2e lrg* approach for 3D reconstruction of the specimen without rotation.

Real measurement data

Experiments

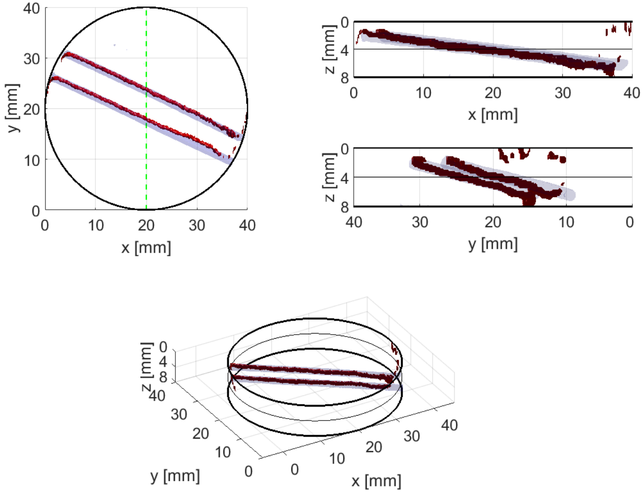


Figure: Using the large end-to-end *e2e lrg* approach for 3D reconstruction of the specimen with a rotation of 25°.

Real measurement data

Experiments

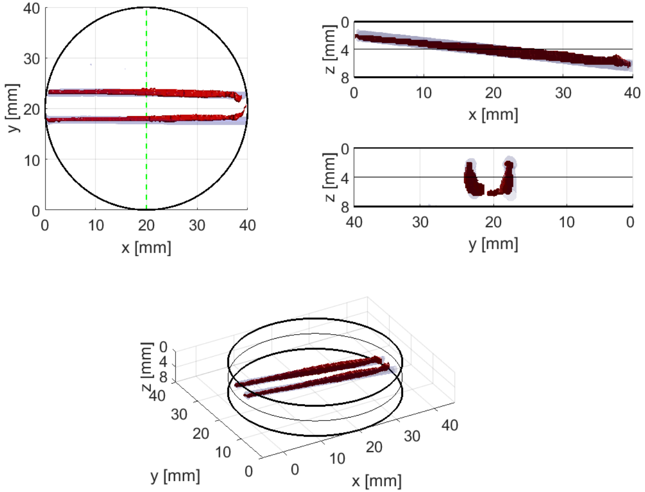
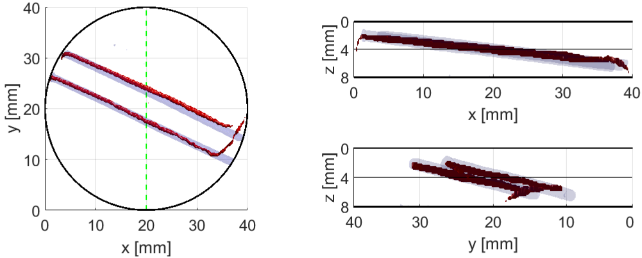


Figure: Using the large hybrid *lrg* approach for 3D reconstruction of the specimen without rotation.



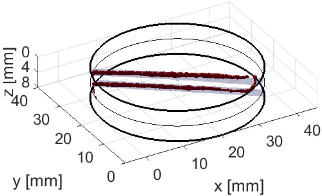


Figure: Using the large hybrid *lrg* approach for 3D reconstruction of the specimen with a rotation of $25°$.

Reformulate the fist step

$$r(\mathbf{v}, \alpha) := \|\mathbf{d} - \mathbf{K}(\alpha)\mathbf{v}\|_2^2 \to \min_{\mathbf{v},\alpha}$$

where

- $\mathbf{d}$ is the surface temperature data,
- $\mathbf{K}$ physics-based forward modeling of the first step,
- $\mathbf{v}$ virtual wave vector.
- α stands for the thermal diffusivity of the material.

Can we define a network to learn α?

- Search the solution to $r(\mathbf{v}, \alpha)$ by Variable Projection (VP).
- Wrap the least-squares estimate $\mathbf{K}^+(\alpha)\mathbf{d}$ into a (VP)layer.
- Define the gradient through α by the theory of VP.

- *Inspirations*
  - CNNs, Wiener-Hammerstein based NNs.
  - Biomedical signal processing (ECG, EEG, EMG, etc.)
  - 1D signal processing with machine learning (neural networks)
- *Traditional approach*
  - Feature extraction + machine learning
  - Time- or frequency domain decomposition
    Fourier transform, Hermite functions, wavelets, statistical descriptors, variable projection (VP), etc.
  - Domain knowledge, model-based methods
  - Explainability
- *Deep Learning*
  - Deep NN, convolutional NN, recurrent NN, etc.
  - Representation learning
- **Idea**
  - Combination: model-based (deep) NN with VP

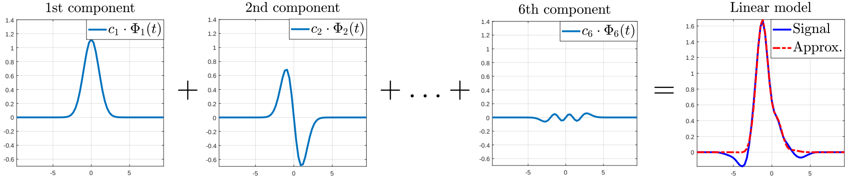
- Linear modeling problem:

$$x \approx \tilde{x} = \sum_{k=0}^{n-1} c_k \Phi_k = \Phi c$$

- Best approximation problem in Hilbert spaces:
  $\mathcal{S} := \mathrm{span}\{\Phi_0, \Phi_1, \ldots, \Phi_{n-1}\} \subset \mathbb{R}^m$ generated subspace

$$\mathrm{dist}(x, \mathcal{S}) := \min_{y \in \mathcal{S}} \|x - y\|_2 = \|x - \tilde{x}\|_2$$

- Solution to the discrete case (linear least squares):
  - Generalized Fourier coefficients: $c = \Phi^+ x$
  - Orthogonal projection: $\tilde{x} = P_{\mathcal{S}} x = \Phi \Phi^+ x$
- Orthogonal transformations with system Φ
  e.g.: trigonometric system, Walsh, rational, and Hermite functions, etc.



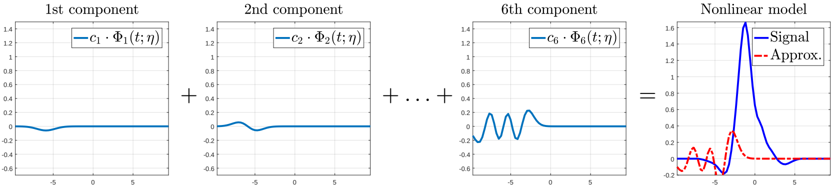
- Nonlinear modeling problem:

$$x \approx \tilde{x} = \sum_{k=0}^{n-1} c_k \Phi_k(\theta) = \Phi(\theta)c$$

where:

- $x \in \mathbb{R}^m$: input data
- $\tilde{x} \in \mathbb{R}^m$: model estimation
- $\Phi_k(\theta) \in \mathbb{R}^m$: parametric function system
- $\Phi(\theta) \in \mathbb{R}^{m \times n}$: system matrix
- c : linear parameters, e.g. $c \in \mathbb{R}^n$ or $c \in \mathbb{C}^n$
- θ : nonlinear system parameters, e.g. $\theta \in \mathbb{D}^p$ for rational functions

- Linear and nonlinear parameters are separated



Nonlinear least-squares approximation of a QRS complex using Hermite functions parametrized by the dilation and the translation.

- General case:

$$r(c,\theta) := \|x - \Phi(\theta)c\|_2^2 \to \min_{c,\theta}$$

- Decomposition:
  - Generalized Fourier coefficients: $c = \Phi^+(\theta)x$
  - Orthogonal projection: $\tilde{x} = P_{\mathcal{S}(\theta)}x = \Phi(\theta)\Phi^+(\theta)x$
  - Variable projection functional [3]:

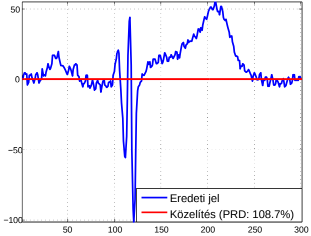
$$r_2(\theta) := \|x - \Phi(\theta)\Phi^+(\theta)x\|_2^2 \to \min_{\theta}$$

- Adaptive transformations
- $\{\Phi_k(\theta) \,|\, 0 \le k < n\}$ adaptive system, $\mathcal{S}(\theta) := \mathrm{span}\{\Phi_k(\theta)\}$
- Function system itself is adapted to the input

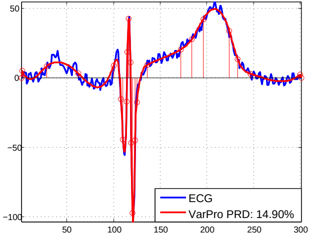
[3] G. H. Golub and V. Pereyra. The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate. *SIAM Journal on Numerical Analysis*, 1973.

Example: B-splines
VPNet

- Nonlinear parameters: free knots $\theta \in \mathbb{R}^p$
- Applications:
  - ECG compression [4]
  - ECG heartbeat classification [5]



(a) Initial estimation.



(b) Optimizing the knots by VP.

[4] P. Kovács and A. M. Fekete. Nonlinear least-squares spline fitting with variable knots. *Applied Mathematics and Computation*, 354:490–501, 2019.

[5] T. Dózsa, G. Bognár, and P. Kovács. Ensemble learning for heartbeat classification using adaptive orthogonal transformations. In *EUROCAST 2019*, Springer LNCS, 2020.

Example: adaptive Hermite functions

VPNet

- Nonlinear parameters: dilation and translation $\theta = [\tau, \lambda]^T \in \mathbb{R}^2$

$$\Phi_k(\tau, \lambda; t) := \sqrt{\lambda} \cdot \Phi_k\left(\lambda(t-\tau)\right) \qquad (t, \tau \in \mathbb{R}, \lambda > 0)$$

- Applications:
  - ECG compression [6]
  - ECG segmentation / delineation [7]
  - ECG, BP, AP waveform modeling [8]
  - ECG heartbeat classification [9]

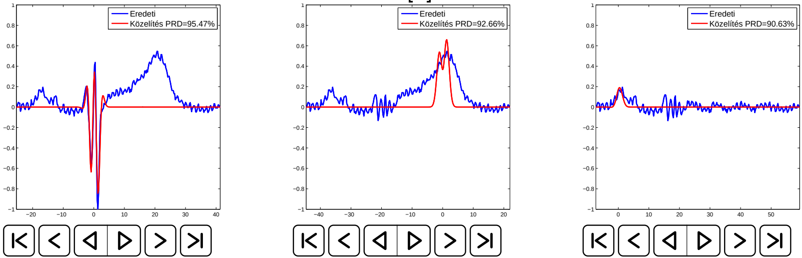
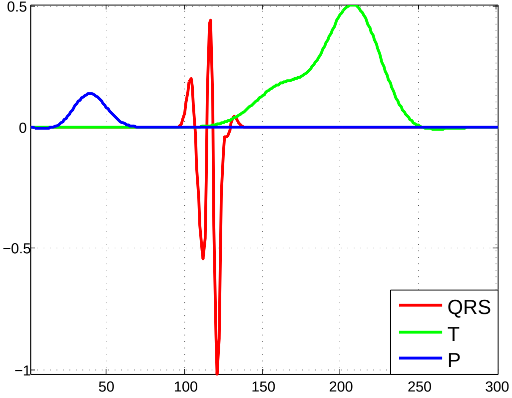


Figure: Fitting the P, T waves, and the QRS complex.

[6] T. Dózsa, P. Kovács. ECG signal compression using adaptive Hermite functions. *Adv Int Syst Comput*, 2015.
[7] P. Kovács, C. Böck, J. Meier, M. Huemer. ECG segmentation using adaptive Hermite functions. In *Asilomar*, 2017.
[8] P. Kovács, C. Böck, T. Dózsa, J. Meier, M. Huemer. Waveform modeling by adaptive weighted Hermite functions. In *ICASSP*, 2019.
[9] T. Dózsa, G. Bognár, P. Kovács. Ensemble learning for heartbeat classification using adaptive orthogonal transformations. In *LNCS*, 2020.

Example: adaptive Hermite (ECG)
VPNet



(d) Segmented heartbeat.

Figure: Segmentation of an ECG based on optimized Hermite functions.

- Nonlinear parameters: inverse poles $\theta \in \mathbb{D}^p$
- Applications:
  - ECG compression [10]
  - ECG segmentation / delineation [11]
  - ECG modeling [12] [13] [14]
  - ECG heartbeat classification [5], [15], [16]
  - EEG seizure detection [17]

[10] P. Kovács, S. Fridli, and F. Schipp. Generalized Rational Variable Projection With Application in ECG Compression. *IEEE Trans Sign Proc*, 2019.
[11] G. Bognár and S. Fridli. ECG Segmentation by Adaptive Rational Transform. In *EUROCAST 2019*, Springer LNCS, 2020.
[12] S. Fridli, P. Kovács, L. Lócsi, and F. Schipp. Rational modeling of multi-lead QRS complexes in ECG signals. *Ann Univ Sci Budapest*, 2012.
[13] S. Fridli, L. Lócsi, and F. Schipp. Rational function system in ECG processing. In *EUROCAST 2011*, Springer LNCS, 2012.
[14] P. Kovács. Rational variable projection methods in ECG signal processing. In *EUROCAST 2017*, Springer LNCS, 2017.
[15] G. Bognár and S. Fridli. Heartbeat Classification of ECG Signals Using Rational Function Systems. In *EUROCAST 2017*, Springer LNCS, 2018.
[16] G. Bognár and S. Fridli. ECG Heartbeat Classification by Means of Variable Rational Projection. *Biomed Sign Process Control*, (to appear)
[17] K. Samiee, P. Kovács, and M. Gabbouj. Epileptic seizure classification of EEG time-series using rational discrete short time Fourier transform. *IEEE Trans Biomed Eng*, 2014.

Example: rational (ECG)
VPNet

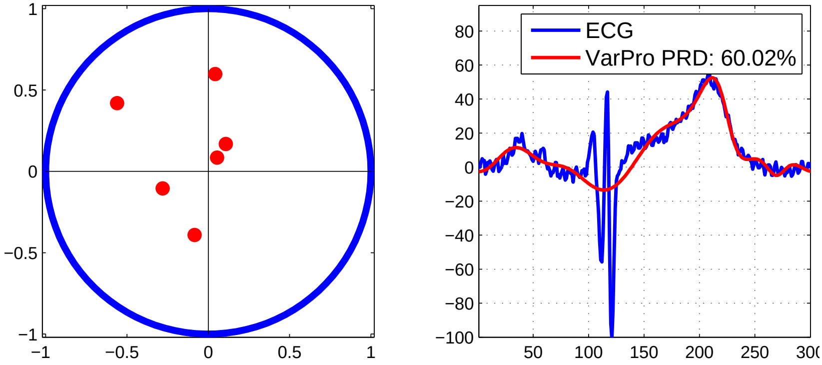
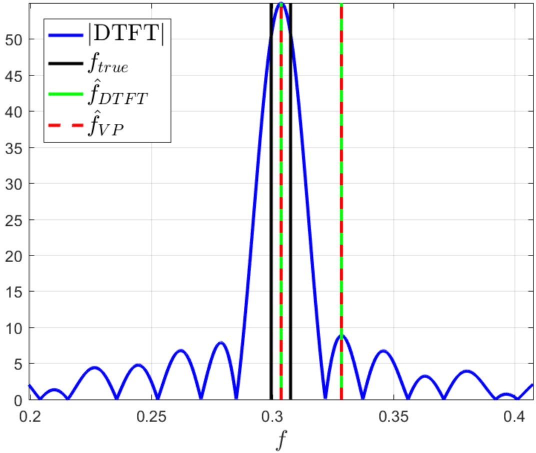


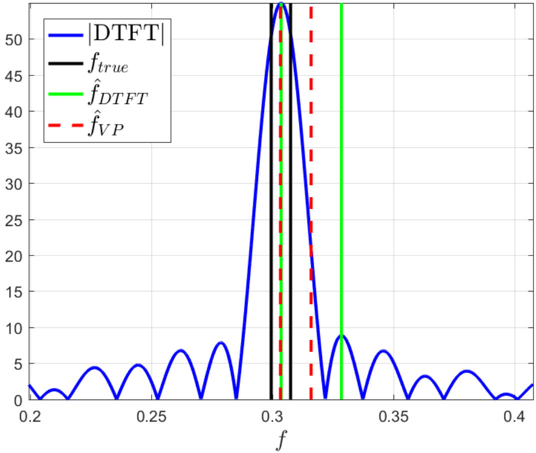
Figure: Example for the rational VarPro algorithm approximating a real ECG from PhysioNet MIT-BIH Arrhythmia Database.



Identification of close frequency components [18].

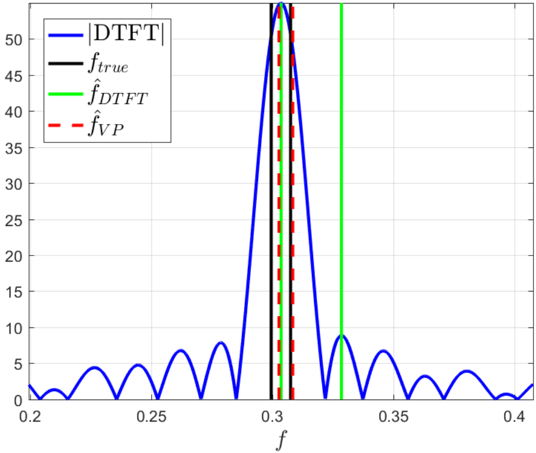
[18] Yuneisy, E. G. G., Kovács, P., Huemer, M., Variable Projection for Multiple Frequency Estimation, in *ICASSP*, 2020, pp. 4811-4815.

Example: trigonometric functions
VPNet



Identification of close frequency components [18].

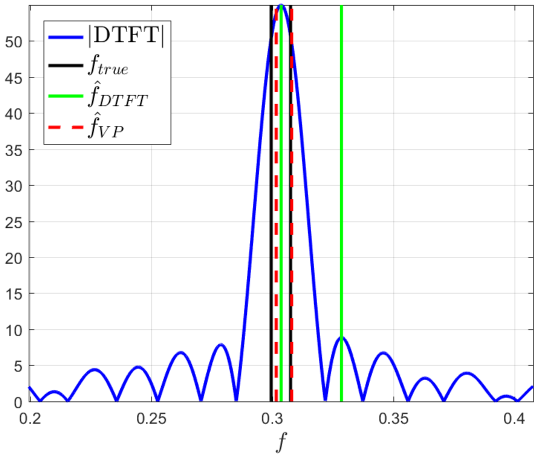
Example: trigonometric functions
VPNet



Identification of close frequency components [18].

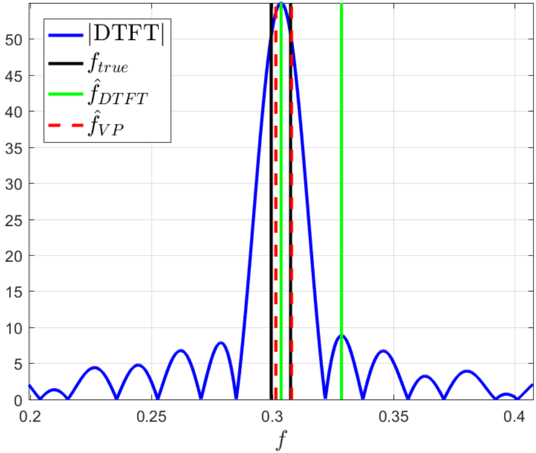
Example: trigonometric functions

VPNet

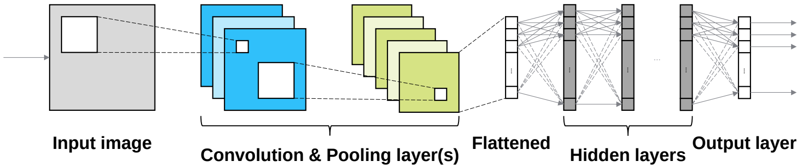


Identification of close frequency components [18].

Example: trigonometric functions
VPNet

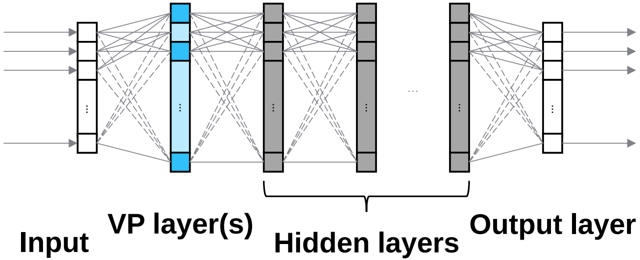


Identification of close frequency components [18].



- Convolutional layers: convolution with nonlinear activation
- Pooling layers: dimension reduction
- Representation learning: built-in multilevel feature extraction
- Input: raw or preprocessed image
- Note: 1D CNN [19]

[19] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman. 1D convolutional neural networks and applications: A survey, 2019.



- Input: raw or preprocessed signal
- VP layer(s): projection of the form

$$x \mapsto f^{(\mathsf{vp})}(x) = \Phi^{+}(\theta)x = c \qquad \text{(classification)}$$

or

$$x \mapsto f^{(\mathsf{vp})}(x) = \Phi(\theta)\Phi^{+}(\theta)x = \tilde{x} \qquad \text{(regression)}$$

- *Novelty*
  - novel model-driven network architecture
  - application: 1D signal processing
- *Generality*
  - arbitrary parameterized function systems
  - domain knowledge
- *Interpretability*
  - built-in feature extraction
  - interpretable parameters: nonlinear VP system parameters
  - direct connection with morphological properties
- *Simplicity*
  - few system parameters only
  - compact architecture (cf. CNN and DNN)

- Offline supervised learning
- Backpropagation, stochastic gradient descent
- Gradients of VP coefficients:

$$f^{(\mathsf{vp})}(x) = \Phi^+(\theta)x, \qquad \frac{\partial f^{(\mathsf{vp})}}{\partial \theta_j} = \frac{\partial \Phi^+(\theta)}{\partial \theta_j}x,$$

where [3]

$$\partial \Phi^+ = -\Phi^+ \partial \Phi \Phi^+ + \Phi^+ \left[\Phi^+\right]^T \partial \Phi^T (I - \Phi\Phi^+) + \\ +(I - \Phi^+\Phi)\partial \Phi^T \left[\Phi^+\right]^T \Phi^+$$

- Gradients of VP projection:

$$f^{(\mathsf{vp})}(x) = \Phi(\theta)\Phi^+(\theta)x, \qquad \frac{\partial f^{(\mathsf{vp})}}{\partial \theta_j} = \frac{\partial (\Phi(\theta)\Phi^+(\theta))}{\partial \theta_j}x,$$

where [3]

$$\partial(\Phi\Phi^+) = (I - \Phi\Phi^+)\partial\Phi\Phi^+ + \left((I - \Phi\Phi^+)\partial\Phi\Phi^+\right)^T$$

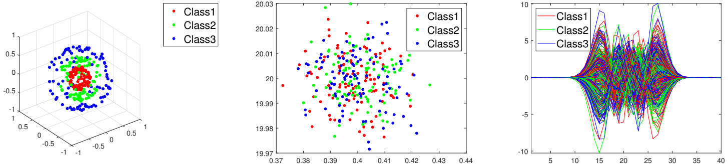
- Implementation: PyTorch / native NumPy framework
  (custom plugin / own implementation)
- Function system: adaptive Hermite functions

$$\Phi_k(\tau, \lambda; x) := \sqrt{\lambda} \cdot \psi_k\left(\lambda(x - \tau)\right) \qquad (x, \tau \in \mathbb{R},\ \lambda > 0)$$

  (nonlinear parameters: translation and dilation)
- Synthetic dataset generation
- Real-world dataset: MIT-BIH Arrhythmia Database
  (ECG classification problems)
- Exhausting evaluation of hyperparameters
- Comparison with fully-connected (FCNN) and convolutional neural networks (CNN)

Synthetic Hermite dataset

Experiments



(a) Coefficients (b) System parameters (c) Samples

- Samples: linear combinations of Hermite functions of the form

$$x_k = \Phi(\tau_k, \lambda_k) \cdot c_k.$$

- Separable coefficients (3 classes)

  $(c_{k,0}, c_{k,1}, c_{k,2}) \in \mathbb{R}^3$: *on spherical shells by classes*

  $c_{k,3}$ *and* $c_{k,4}$: *amplitude normalization*

- Similar system parameters

  τ_k *and* λ_k: *generated randomly with given mean and variance*

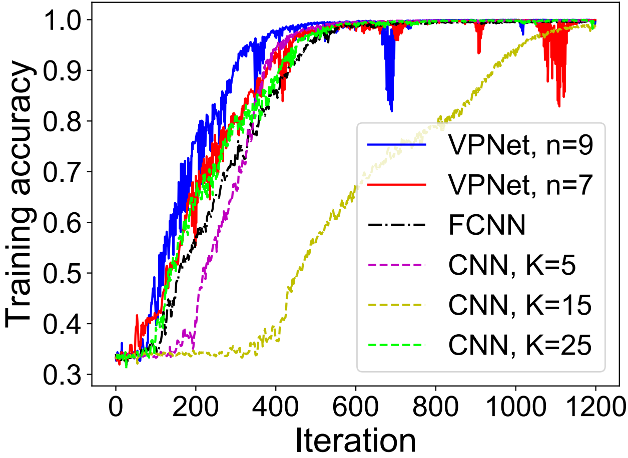


Figure: Best training curves

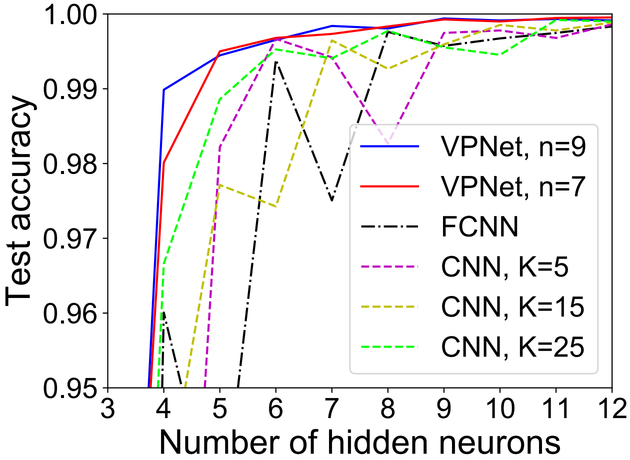
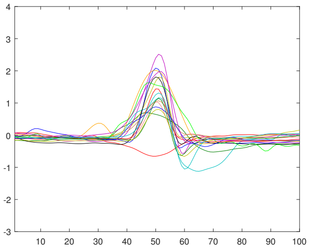
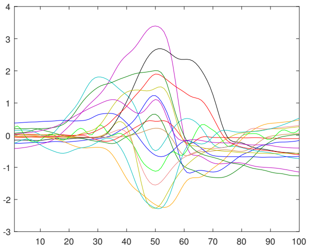


Figure: Best test accuracy depending on the number of hidden neurons

Real ECG dataset
Experiments



(a) Normal beats



(b) Ventricular ectopic beats

- PhysioNet MIT-BIH Arrhythmia Database
- Reduced, balanced subset
- Normal $\leftrightarrow$ ventricular ectopic heartbeats
- Training: 4260-4260 beats (DS1), test: 3220-3220 (DS2)

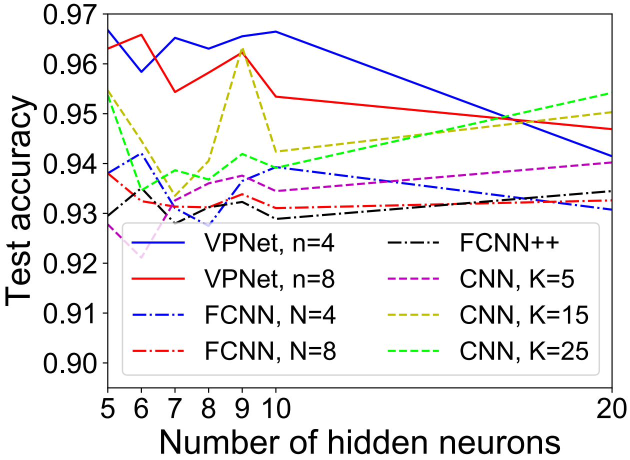


Figure: Best test accuracy depending on the number of hidden neurons

- *Summary*
  - Novel model-based architecture for 1D signal processing
  - General, flexible construction
  - Compactness
  - Explainability, interpretable parameters
  - Preliminary results: outperforms FCNN and CNN wrt. convergence and accuracy
- *Further research*
  - Mathematical and computational properties
  - New fields of applications
  - Classification, regression, clustering problems
  - Different architectures, other ML methods combined with VP
- *Cooperation partners*

JOHANNES KEPLER UNIVERSITY LINZ

INSTITUTE OF SIGNAL PROCESSING

SAL SILICON AUSTRIA LABS