Grad-CAM-Inspired Interpretation of Nearfield Acoustic Holography using Physics-Informed Explainable Neural Network

Hagar Kafri [1], Marco Olivieri [2], Fabio Antonacci [2], Mordehay Moradi [1], Augusto Sarti [2], Sharon Gannot [1]

[1] Bar-Ilan University, Ramat-Gan, Israel

[2] Politecnico di Milano, Italy - Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB)

ISPL Image and Sound Processing Lab

POLITECNICO MILANO 1863

CONTEXT

ISSUE: Visual explanation of a recent physics-informed approach for Nearfield Acoustic Holography (NAH) using vibrating rectangular plates with different boundary conditions and violin top plates with complex shapes.

GOAL: Interpret and explain the decision-making process of the Kirchhoff-Helmholtz-based Convolutional Neural Network (KHCNN) architecture.

PROPOSED METHODOLOGY: Propose a Grad-CAM-inspired approach for the visual explanation of neural network architecture for regression problems.

ABSTRACT

The interpretation and explanation of decision-making processes of neural networks are becoming a key factor in the deep learning field. Although several approaches have been presented for classification problems, the application to regression models needs to be further investigated. In this manuscript we **propose a Grad-CAM-inspired approach** for **the visual explanation** of **neural network architecture for regression problems**. We apply this methodology to a recent physics-informed approach for Nearfield Acoustic Holography, called Kirchhoff-Helmholtz-based Convolutional Neural Network (**KHCNN**) architecture.

We focus on the interpretation of KHCNN using vibrating rectangular plates with different boundary conditions and violin top plates with complex shapes. **Results highlight the more informative regions of the input** that **the network exploits to correctly predict the desired output**. The **devised approach** has been **validated** in terms of **NCC and NMSE** using the original input and the filtered one coming from the algorithm.

1. SCIENTIFIC BACKGROUND - Nearfield Acoustic Holography

The complex exterior radiated pressure (i.e., magnitude and phase information) measured at a point r due to the vibration of the structure at $\omega = 2\pi f$ can be **formulated with the Kirchhoff-Helmholtz (KH) integral** as:

$$p(\mathbf{r},\omega) = \int_S p(\mathbf{s},\omega)\frac{\partial}{\partial \mathbf{n}} g_\omega(\mathbf{r},\mathbf{s}) d\mathcal{S} - j\omega\rho_0 \int_S v_n(\mathbf{s},\omega) g_\omega(\mathbf{r},\mathbf{s}) d\mathcal{S}$$

- s a point of surface S
- ρ_0 the density of the medium (for air $\rho_0 = 1.225\ kg\ m^{-3}$)
- n the outward vector normal to the surface at s
- $v_n(s,\omega)$ the normal velocity field
- $p(s,\omega)$ pressure field on the vibrating surface considering the propagation from s to r
- $g_\omega(r,s)$ Free-field Green's function

Nearfield Acoustic Holography (NAH) aims at computing $v_n(s,\omega)$ starting from $p(r,\omega)$ acquired by a microphone array on the holographic plane $\mathcal{H}$ ($r \in \mathcal{H}$). Therefore, the goal of NAH relies on the **inversion of the equation** above, namely:

$$\hat{v}_n(\mathbf{s},\omega)\Big|_{\mathbf{s}\in\mathcal{S}} \approx \Gamma^{-1}\big[p(\mathbf{r},\omega)\big]\Big|_{\mathbf{r}\in\mathcal{H}}$$

2. SCIENTIFIC BACKGROUND - Grad-CAM

Gradient-weighted Class Activation Mapping (Grad-CAM) is a **post-hoc explanation via visualisation** of class discriminative activation for a network. **Grad-CAM takes advantage on a key property of deep convolutional layers.**

Grad-CAM determines the neuron importance weights as the global average pooling of the gradients over the spatial dimension, namely:

$$\alpha_k^c = \frac{1}{Z}\sum_i^H \sum_j^W \frac{\partial y^c}{\partial A_{ij}^k}$$

- c a class $\in C$
- k feature maps in the last convolution layer (A^k)
- y^c the score for class c
- α_k^c the neuron importance weights

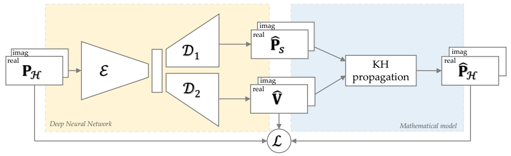
$$\mathbf{R}^c = \mathrm{ReLU}\left(\sum_k \alpha_k^c \mathbf{A}^k\right)$$

$Z = H \times W$ the spatial resolution of the feature map with height H and width W.

Applying **ReLU** function to **the linear weighted sum of the activation maps.**

3. PROPOSED METHOD

The block scheme of KHCNN architecture is:



- **INPUT**: $P_{H(\omega)} \in \mathbb{C}^{N_1 \times N_2}$ - the normal velocity field and the pressure field on S
- **OUTPUT**: $\hat{P}_{H(\omega)} \in \mathbb{C}^{M_1 \times M_2}$ - the acoustic pressure fields

NETWORK ARCHITECTURE:

- *Deep Neural Network Block* - CNN with one encoder E and two decoders D_1 and D_2 to estimate the latent variable $\hat{P}_s$ and the velocity $\hat{V}$ from the input.
- *Mathematical model block* – contain layers of mathematical steps according to the KH integral to create the estimate $\hat{P}_H$ for the training.

THE LOSS FUNCTION:

The proposed algorithm aims to visualize the heatmap **H to highlight the important regions** of the input P_H needed to produce the KHCNN estimate $\hat{V}$.

We use **the loss function** defined as: $\mathcal{L} = 0.5 \cdot \mathcal{L}_{Im} + 0.5 \cdot \mathcal{L}_{Re}$, with:

$$\mathcal{L}_{\mathrm{Re}} = \left\|\mathrm{Re}(\mathbf{V}) - \mathrm{Re}\left(\hat{\mathbf{V}}\right)\right\|_2^2 + \left\|\mathrm{Re}(\mathbf{P}_\mathcal{H}) - \mathrm{Re}\left(\hat{\mathbf{P}}_\mathcal{H}\right)\right\|_2^2,$$

$$\mathcal{L}_{\mathrm{Im}} = \left\|\mathrm{Im}(\mathbf{V}) - \mathrm{Im}\left(\hat{\mathbf{V}}\right)\right\|_2^2 + \left\|\mathrm{Im}(\mathbf{P}_\mathcal{H}) - \mathrm{Im}\left(\hat{\mathbf{P}}_\mathcal{H}\right)\right\|_2^2,$$

For regression problem, the Grad-Cam weights are modified:

$$\alpha_k = \frac{1}{Z}\sum_i \sum_j \frac{\partial \mathcal{L}}{\partial A_{ij}^k}$$

4. EVALUATION METRICS

- Normalized Cross Correlation (**NCC**)
- Normalized Mean Square Error (**NMSE**)

$$\mathrm{NCC} = \frac{|\hat{\mathbf{v}}_f^H \cdot \hat{\mathbf{v}}|}{\|\hat{\mathbf{v}}_f\|_2 \cdot \|\hat{\mathbf{v}}\|_2}, \quad \mathrm{NMSE} = 10\log_{10}\left(\frac{\mathbf{e}^H \cdot \mathbf{e}}{\hat{\mathbf{v}}^H \cdot \hat{\mathbf{v}}}\right)$$

between the KHCNN estimates coming from the acquired input pressure and the filtered one by the resulting heatmap H

5. SETUP

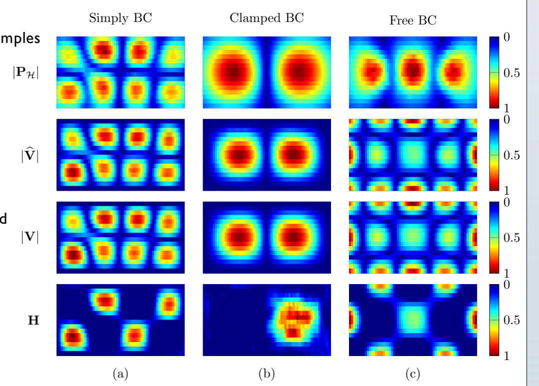
Using the pre-trained KHCNN architecture, we focus on two datasets:

- 672 aluminum **rectangular** plates (different dimensions and **boundary conditions**)
- 1568 **violin** top plates made of Sitka spruce (**different resolutions** and only free BC)
- The datasets were generated using COMSOL Multiphysics® software simulating the radiated pressure in M points and the normal velocity field in N points of different plates excited at different ω.

We consider high-resolution input with $M = 16\times64$ points and low-resolution input with $M = 8\times8$ points. In both cases KHCNN estimates the desired velocity in $N = 16\times64$ points.

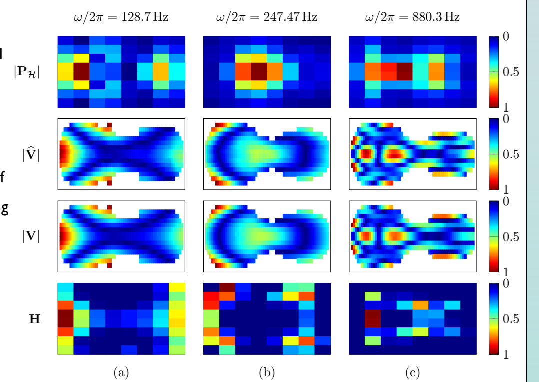
6. RESULTS

The fig shows three examples of rectangular plates having different BCs along with the input hologram pressure in $M = 1024$ points, the velocity estimate and its ground truth in $N = 1024$ points. From the output of decoder D_2 we applied the algorithm to obtain the heatmap H. In general, inspecting the resulting heatmaps H, we noticed:



- the network focuses on the regions where vibrational **lobes** of the velocity fields are present.
- KHCNN exploits **the symmetry of vibrations** and **the edges** for different BCs inferring the estimates as interpolation of the vibrational patterns.

For the violin top plate we analyze the KHCNN architecture using $M = 64$ points for the input pressure. Again, three examples of violin top plates vibrating at different f with the computed $H \in \mathbb{R}^{8\times8}$.



We can notice that the **network is more affected by the edges of the input** mainly at low f. Conversely, when f increases the network focuses on the internal regions.

References

[1] M. Olivieri, R. Malvermi, M. Pezzoli, M. Zanoni, S. Gonzalez, F. Antonacci and A. Sarti, "Audio information retrieval and musical acoustics", IEEE Instrumentation & Measurement Magazine, vol. 24, no. 7, pp. 10-20, 2021.

[2] M. Olivieri, M. Pezzoli, R. Malvermi, F. Antonacci and A. Sarti, "Near-field acoustic holography analysis with convolutional neural networks", INTER-NOISE and NOISE-CON Congress, vol. 261, pp. 5607-5618, August 2020.

[3] M. Olivieri, M. Pezzoli, F. Antonacci and A. Sarti, "Near field acoustic holography on arbitrary shapes using convolutional neural network", European Signal Processing Conference, August 2021.

[4] M. Olivieri, M. Pezzoli, F. Antonacci and A. Sarti, "A physics-informed neural network approach for nearfield acoustic holography", Sensors, vol. 21, no. 23, pp. 7834, 2021.

[5] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization", Proceedings of the IEEE on computer vision, pp. 618-626, 2017