

#97



Linear systems theoretic approach to interpretation of spatial and temporal weights incompact CNNs: Monte-Carlo study.

Artur Petrosyan¹, Mikhail Lebedev¹, Alexey Ossadtchi¹ (aossadtchi@hse.ru).

¹. Center for Bioelectric Interfaces National Research University Higher School of Economics, Moscow, Russia

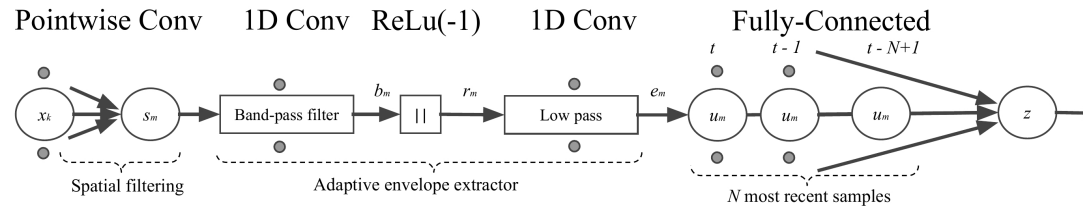


ABSTRACT

Interpretation of the neural networks architectures for de-coding the signals of the brain usually reduced to the analysis of temporal and spatial weights. We describe a theoretically justified method of their interpretation within the simple architecture based on a prior knowledge of the subject area. This architecture is comparable in decoding quality to the winner of the BCI IV competition and allows for automatic engineering of physiologically meaningful features. To demonstrate the operation of the algorithm, we performed Monte Carlo simulations and received a significant improvement in the restoration of patterns for different noise levels and also investigated the relation between the decoding quality and patterns reconstruction fidelity

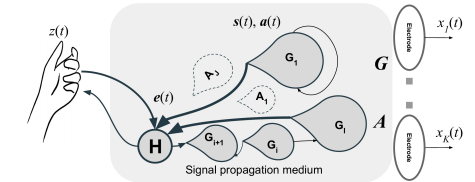
NEURAL NETWORK ARCHITECTURE

Adaptable and compact CNN architecture based on the idea of described phenomenological model. It consists of spatial filtering, adaptive envelope extractor, and a fully-connected layer. Spatial filtering is done via a pointwise convolution layer used to unmix the sources. Adaptive envelope extractor is in a form of two depth-wise convolutions for bandpass and lowpass filtering, used with non-trainable batch normalization (for explicit power extraction) and absolute value nonlinearity in-between. The fully-connected layer is used in order to model envelope to kinematics transformation as a function of lagged envelopes of layers signals extracted previously.



PHENOMENOLOGICAL MODEL

Neural populations G1 - G1, which are responsible for movement, generate activity $e(t)$ that is further translated into a trajectory $z(t)$ of movements with some non-linear transform H. We assume that there are populations SA_1-A_JS with unrelated movement activity. Its activity is mixed into the sensors. At each time step we observe $x(t)$ - vector of sensor signals



PERFORMANCE

Correlation between true and predicted kinematics of the winning solution for BCI competition IV dataset (Winner) and proposed architecture (NET)

Subject	1[2]	Thumb	Index	Middle	Ring	Little
Winner	.58 .51 .69	.71 .37 .46	.14 .24 .58	.53 .47 .58	.29 .35 .63	
NET	.54 .50 .71	.70 .36 .48	.20 .22 .50	.58 .40 .52	.25 .23 .61	

CONCLUSIONS

1. We proposed a theoretically justified method for the interpretation of temporal and spatial weights of the CNN architecture composed of simple envelope extractors.
2. This result extends already existing approaches to weights interpretation.
3. With Monte-Carlo simulations we were able to demonstrate that the proposed approach accurately recovers both spatial and temporal patterns of the underlying phenomenological model for a broad range of signal to noise ratio values.

ACKNOWLEDGMENTS

This work is supported by the Center for Bioelectric Interfaces NRU HSE, RF Government grant, ag. No. 14.641.31.0003.

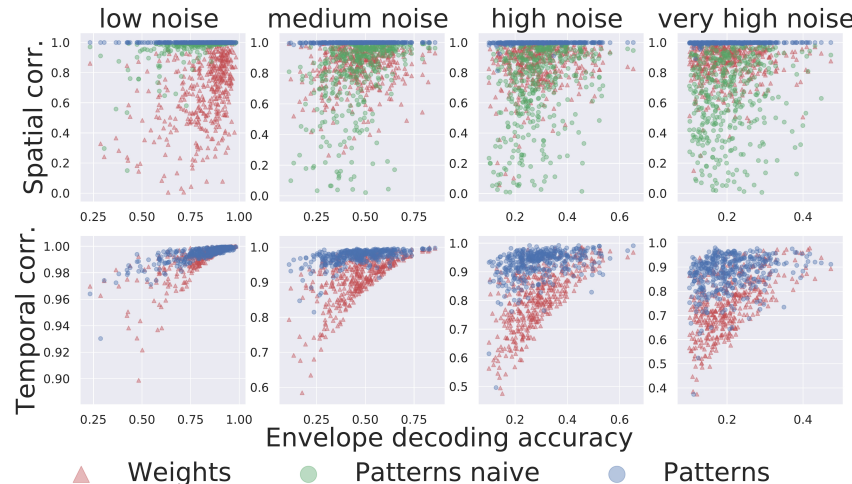
WEIGHTS INTERPRETATION

Assume that the data are processed in chunks of size N equal to the length of the temporal convolutional layer weights h_m . $X(t) = [x(t), x(t-1), \dots, x(t-N+1)]$. Since the set of envelopes maps isomorphically onto a set of analytic signal, the task of tuning the weights of the first three layers of our architecture to predict envelopes $e_m(t)$ can be replaced with a regression problem of learning and correcting spatial and temporal weights to get the analytic signal $b_m(t)$ giving rise to the envelope. Assume that the temporal weights are fixed to their optimum value h_m , then the optimal spatial filter weights can be obtained as:

$w_m^* = \arg \min_{w_m} \{ \|b_m(n) - w_m^T X(t) h_m^*\|_2^2 \}$ and therefore assuming statistical independence of the rhythmic LFPs $\{h_m(t)\}$, $m=1, \dots, M$ the spatial pattern of the underlying neuronal population is

$g_m = E\{Y(t)Y(t)^T\} w_m^* = R_m^Y w_m^*$
Symmetrically we can write an expression for the temporal weights interpretation as
 $q_m = E\{V(t)V(t)^T\} h_m^* = R_m^V h_m^*$

MONTE-CARLO SIMULATIONS



We performed Monte-Carlo study with different spatial configuration of sources at each trial. For each realisation of the generated data we have trained the DNN to predict the kinematic variable and then computed the patterns. Figure shows that only the spatial Patterns interpreted using branch-specific temporal filters match well the simulated topographies of the true underlying sources. The spectral patterns recovered using the proposed approach also appear to match well with the true spectral profiles of the underlying sources, while directly considering the Fourier coefficients of the temporal convolution layer weights results into erroneous spectral profiles.