

Artificial Intelligence and Deep Learning

at BME TMIT SmartLab

BME, Budapest University of Technology and Economics, Hungary
since 1782, 8 faculties, app. 21.000 students

**TMIT (Dept. of Telecommunications and Media Informatics,
since 1949), ~80 employees**

1

SmartLab @ BME TMIT (since 1978)
Head of Laboratory: Dr. Géza Németh

~20 employees

2 DSc, 9 PhD

5+1 PhD students

20-25 MSc & BSc students



BME TMIT SmartLab

Speech Communications and Smart Interactions (since 1978, from research ideas to real-life applications)

Industrial partners: T-Mobile, Vodafone, NCR, Wincor-Nixdorf, Hungarian Railways, E-Net, OTP Bank, ...

Spinoffs: Aitia Inc, SpeechTex Ltd, ThinkTech Ltd, ...

Link to IoT and Industry 4.0: Arrowhead framework and MANTIS project

Deep Learning R&D&E

- 6 **NVidia GPU Research Grants** in the past 2 years
- July 2016: official **NVidia GPU Education Center**
(the only one in Hungary)
- September 2017: Gyires-Tóth Bálint, **Deep Learning Institute (DLI) Instructor and University Ambassador**



<http://smartlab.tmit.bme.hu>

SmartLab
Intelligent Interactions



AI & Deep Learning related research directions, results

Applied AI & Deep Learning Research

- Audio and speech generation, recognition and classification
- Image recognition, object detection and segmentation
- Modeling sensor data
- Optimizing deep models for low-resource devices
- Natural Language Processing (NLP)
- Smart devices (phones, TVs, robots, vehicles, ...)

Fundamental AI & Deep Learning Research

- Modeling really long-term dependencies
- Modeling rare events
- Q-Learning (Reinforcement Learning)



Proposed SmartLab contributions to the AI Platform

Resources

- Industrial grade technologies and engines
- 40 years of AI related R&D (from research ideas to real-life applications)
- Industry/end-user partnerships
- Databases

To research challenges

- Adaptation to human dialogue partners by (social) robots (e.g. digital baby, context dependent interaction handling)
- Long-term interaction memory and planning

Application challenges

- Smart Interactions in
 - vehicles
 - manufacturing and industrial scenarios
 - health & robotics



<http://smartlab.tmit.bme.hu>

SmartLab
Intelligent Interactions



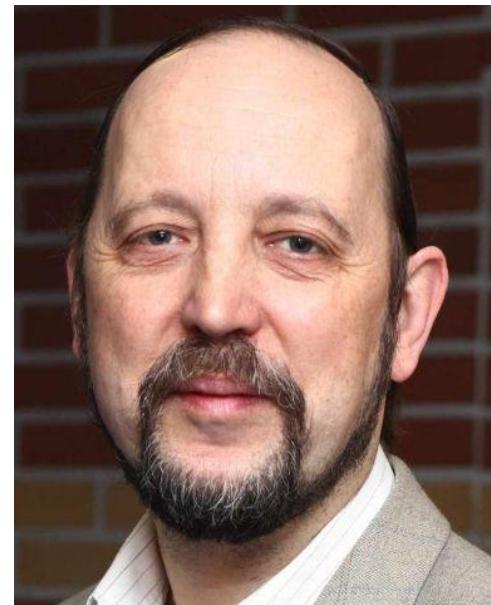
This is the end of the 1 minute presentation

If you found it interesting, please read the rest,
and contact me here or later.

E-mail: nemeth@tmit.bme.hu

Mobile/Viber: +36 30 230 2940

Skype: nemeth_geza



<http://smartlab.tmit.bme.hu>

SmartLab
Intelligent Interactions



Audio Classification with Deep Learning

Acoustic analysis of traffic environment

E.g. classification of human speech, heavy vehicles, tram, ambulance siren, etc.

Several hours of labeled 4 channel acoustic signals.

Application domain: event detection inside and outside the vehicle

Large-Scale Bird Song Classification in Noisy Environment (BIRDCLEF)

<https://github.com/bapalto/birdsong-keras>



<http://smartlab.tmit.bme.hu>

SmartLab
Intelligent Interactions

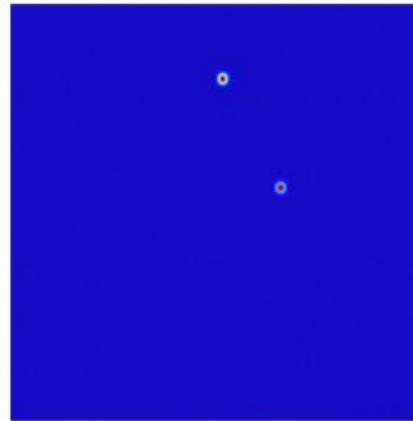
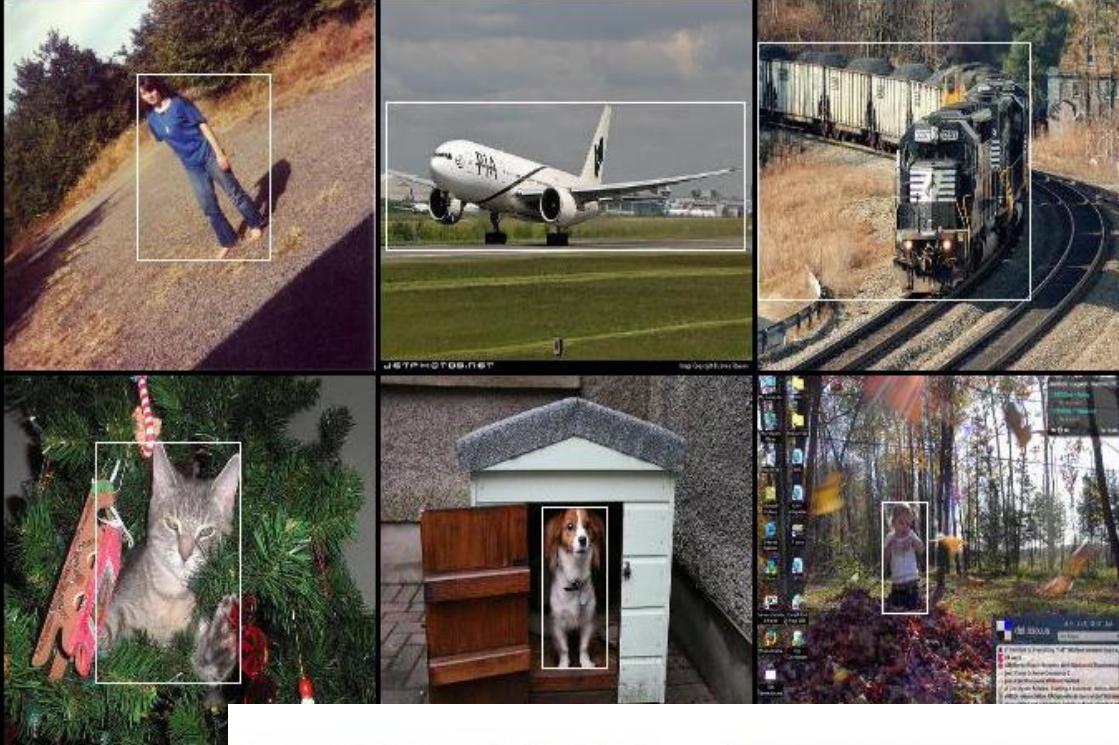


Image recognition, object detection and segmentation

Using state-of-the-art deep neural network models (including Inception v3+, Xception, VGG-16, VGG-19, etc.).

Application domains:

- Unsupervised image classification,
- Object detection and segmentation,
- Sport analytics – pose estimation,
- IMAGECLEF: large-scale image recognition challenge.



Modeling Sensor Data with Deep Learning

1D convolution, LSTM layers

Residual, highway, dense architectures

Stacked models

Application domains:

- User and vehicle identification and authentication,
- Activity and behaviour analysis,
- Detecting emergency situations,
- Early detection of failures in vehicles and manufacturing systems.



<http://smartlab.tmit.bme.hu>

SmartLab
Intelligent Interactions



Modeling Sensor Data with Deep Learning



M Ú E G Y E T E M 1 7 8 2

<http://smartlab.tmit.bme.hu>

SmartLab
Intelligent Interactions



Optimizing Deep Learning for Low-Resource Devices

Inference on low-resource devices.

Performance, power and memory consumption optimization.

Model optimization:

- Undersampling weights
- Compressing weights

Android and iPhone
implementations.



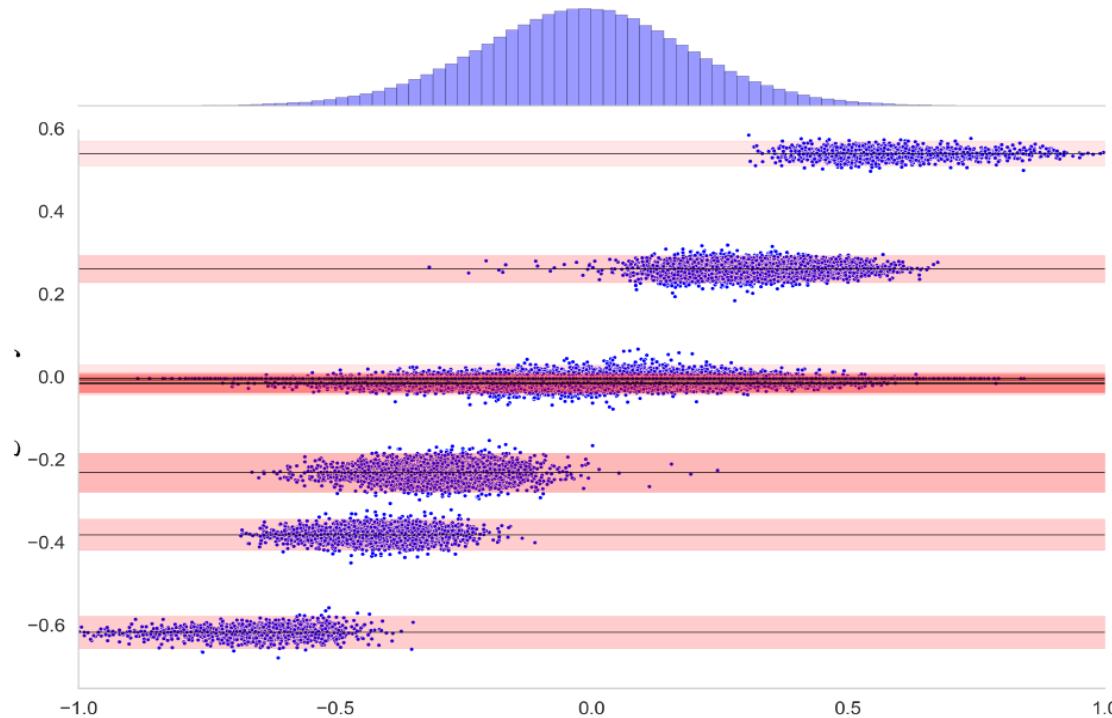
MŰEGYETEM 1782

<http://smartlab.tmit.bme.hu>

SmartLab
Intelligent Interactions



Optimizing Deep Learning for Low-Resource Devices



2.27MB → 78 KB (~29x)

Speech Technologies with Deep Learning

Text-To-Speech, Speech-To-Text, sentiment analysis (e.g. depression, speech disorder, throat cancer), speech as biosignal
Parametric approach and WaveNet

Application domains:

- Railway station announcements,
- On-board vehicle announcement systems,
- Real-time, automatic TV live subtitling,
- Call center automation,
- Social robotics
- Mobile apps (e.g. SMS and e-mail reading, talking manual)



<http://smartlab.tmit.bme.hu>

SmartLab
Intelligent Interactions



WaveNet: raw audio regression

- Google, September 2016
(later: Deep Voice, Baidu, February 2017)
- Idea comes from PixelCNN / PixelRNN
- Novel approach: model raw audio waveform



1 Second

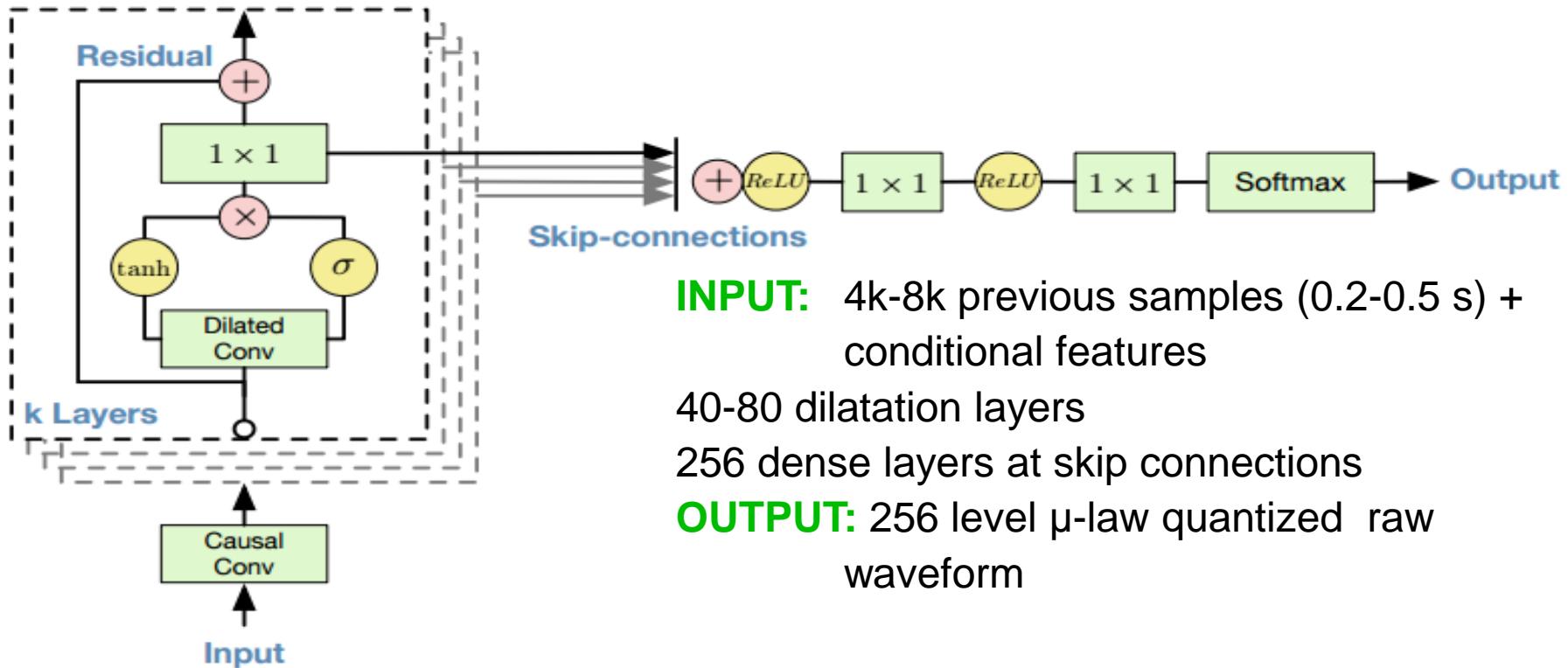


<http://smartlab.tmit.bme.hu>

SmartLab
Intelligent Interactions



WaveNet: raw audio regression



Deep Learning based Natural Language Processing

1D convolution and LSTM based language models (including BOW, CBOW, seq2seq, Tree LSTM, 1Dconv).

Application domains:

- Sentiment analysis
 - Twitter, Reddit, etc. feeds
 - News portal
- End-To-End Speech Synthesis and Recognition systems
- Ensembles with time series



Deep Learning Projects in Initial Phase

- Demand forecast in Multinational Environment
- Attacking deep neural network based image recognition systems



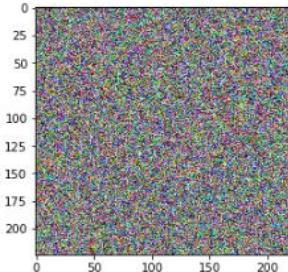
<http://smartlab.tmit.bme.hu>

SmartLab
Intelligent Interactions



Deep Learning Projects

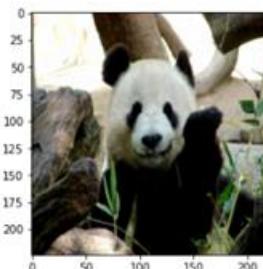
Attacking Deep Neural Nets



```
295 # 0.986627 # American black bear, black bear  
297 # 0.00371121 # sloth bear, Melursus ursinus  
294 # 0.00114846 # brown bear, bruin, Ursus arctos  
342 # 0.000920139 # wild boar, boar, Sus scrofa  
106 # 0.000625546 # wombat
```



```
348 # 0.939152 # ram, tup  
349 # 0.0608187 # bighorn, bighorn sheep, cimarron  
355 # 1.47454e-05 # llama  
350 # 1.26775e-05 # ibex, Capra ibex  
181 # 3.52229e-07 # Bedlington terrier
```



```
68 # 0.507021 # sidewinder, horned rattlesnake  
66 # 0.16484 # horned viper, cerastes, sand viper  
63 # 0.0872896 # Indian cobra, Naja naja  
67 # 0.0366615 # diamondback, diamondback rattlesnake  
359 # 0.0296715 # black-footed ferret, ferret
```

Some recent publications

1. Zainkó Csaba, Tóth Bálint Pál, Németh Géza, Magyar nyelvű WaveNet kísérletek, In: XIII. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY2017), Szeged, Magyarország, 2017, pp. 205-216
2. Péter Nagy, Géza Németh, DNN-Based Duration Modeling for Synthesizing Short Sentences, In: Speech and Computer, Budapest, Magyarország, 2016, pp. 254-261
3. Tóth Bálint Pál, Neurális hálózatok: Beszélő számítógépek mély gondolatokkal, In: ÉLET ÉS TUDOMÁNY, vol. 30, no. LXXI, 2016, p. 3
4. Tóth Bálint Pál, Czeba Bálint, Convolutional Neural Networks for Large-Scale Bird Song Classification in Noisy Environment, In: Working Notes of Conference and Labs of the Evaluation Forum, Évora, Portugália, 2016, p. 8 link
5. Bálint Pál Tóth, Balázs Szórádi, Géza Németh, Improvements to Prosodic Variation in Long Short-Term Memory Based Intonation Models Using Random Forest, In: Speech and Computer, Budapest, Magyarország, 2016, pp. 386-394 Scopus DOI WoS
6. Bálint Pál Tóth, Kornél István Kis, György Szaszák, Géza Németh, Ensemble Deep Neural Network Based Waveform-Driven Stress Model for Speech Synthesis, In: 18th International Conference on Speech and Computer SPECOM 2016, Budapest, Magyarország, 2016, p. 8 Scopus DOI
7. Bálint Pál Tóth, Márton Osváth, Dávid Papp, Gábor Szücs, Deep Learning and SVM Classification for Plant Recognition in Content-Based Large Scale Image Retrieval, In: Working Notes of Conference and Labs of the Evaluation Forum, Évora, Portugália, 2016, p. 9 link
8. Bálint Pál Tóth, Tamás Gábor Csapó, Continuous Fundamental Frequency Prediction with Deep Neural Networks, In: European Signal Processing Conference (EUSIPCO 2016), Budapest, Magyarország, 2016, pp. 1348-1352 Scopus DOI pdf

Some recent publications

9. Kálmán Abari, Tamás Gábor Csapó, Bálint Pál Tóth, Gábor Olaszy, From text to formants - indirect model for trajectory prediction based on a multi-speaker parallel speech database, In: Proceedings of the Annual Conference of the International Speech Communication Association (INTERSPEECH 2015), Dresden, Németország, 2015, pp. 623-627
10. Kele Xu, Pierre Roussel, Tamás Gábor Csapó, Bruce Denby, Convolutional neural network-based automatic classification of midsagittal tongue gestural targets using B-mode ultrasound images, In: JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, vol. 141, no. 6, 2017, pp. EL531-EL537
11. Kele Xu, Tamás Gábor Csapó, Pierre Roussel, Bruce Denby, A comparative study on the contour tracking algorithms in ultrasound tongue images with automatic re-initialization, In: JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, vol. 139, no. 5, 2016, pp. EL154-EL160
12. G Kiss, M G Tulics, D Sztahó, A Esposito, K Vicsi, Language independent detection possibilities of depression by speech, In: Esposito A, Faundez-Zanuy M, Esposito A M, Cordasco G, Drugman Th, Solé-Casals J, Morabito F C (szerk.), Recent Advances in Nonlinear Speech Processing. 294 p. , Cham (Németország): Springer International Publishing, 2016. pp. 103-114.
13. Sztahó D, Vicsi K, Estimating the severity of Parkinson's disease using voiced ratio and nonlinear parameters, In: Pavel Král, Carlos Martín-Vide (szerk.), Statistical Language and Speech Processing: 4th International Conference, SLSP 2016, Cham (Svájc): Springer International Publishing, 2016. pp. 96-107. (Lecture Notes in Computer Science; 9918.)
14. Varga A, Tarjan B, Tobler Z, Szaszak G, Fegyo T, Bordas C, Mihajlik P, Automatic Close Captioning for Live Hungarian Television Broadcast Speech: A Fast and Resource-Efficient Approach, LECTURE NOTES IN ARTIFICIAL INTELLIGENCE 9319: pp. 105-112. (2015), 17th International Conference on Speech and Computer (SPECOM). Athens, Greece

Educational activities in AI Project Lab, Thesis work, PhD

- Preparation for project laboratory
- Project Laboratory
- BSc and MSc thesis work
- PhD students

- 6 PhD students
- 20-25 BSc and MSc students



<http://smartlab.tmit.bme.hu>

SmartLab
Intelligent Interactions



Educational activities in AI Deep Learning Symposium



Educational activities in AI AI related BME TMIT SmartLab courses

- 1. Deep Learning in Practice with Python and LUA**
(in Hungarian XY students)
- 2. Human-Computer Interaction**
(in Hungarian XY students and English 47 students)
- 3. Human-Robot Interaction**
(in Hungarian XY students)
- 4. Speech Information Systems**
(in Hungarian 117 students)



<http://smartlab.tmit.bme.hu>

SmartLab
Intelligent Interactions



Educational activities in AI

Deep Learning Class



Educational activities in AI Deep Learning Class

Student feedback:

“I really enjoyed the class, this one I like the most at the university. I liked the modern, agile attitude of the lecturers and that I learned a cutting edge technology in practice.”

“I would say this was the best class I attended at the university. Up-to-date (literally, actual for the lecture’s day) information, a lot of examples and real world problems were solved. I can’t really imagine a better class to dive into deep learning, it gives a good knowledge that can be later enhanced. And one more great thanks for the lecturers, if everything is fine I get my deep learning and machine learning related job on Friday :)”



Contacts

Email: smartlab@tmit.bme.hu

Thank you for your attention!

