GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Seminar on the Physics
of Deep Learning

Using Convolutional Neural Networks for Speech Recognition

Authors: Jan Cammann, Jorge de Heuvel, Marvin Klingner, David Schlegel
(Dated: August 16, 2017)

ADDITIONAL GRAPHICS

2	AguarunaAwajun	48	Galela	93	Lobi	131	Pijin
3	Akha	50	Garhwali	94	Luwo	134	Qeqchi
4	Akoose	51	Gidar	95	Madi	135	QuechuaEastern
5	Alangan	52	Guahibo	97	MalagasyPlateau	136	QuechuaMargosYarowilcaLauricocha
6	AlbanianTosk	54	Hadiyya	99	Marba_Azumeina	137	QuechuaNorthJunin
7	Alladian	55	Hiligaynon	100	MayaMopan	138	QuechuaSouthernConchucosAncash
8	AltaiSouthern	56	Hindi	101	MazahuaCentral	140	Romanian
9	Anufo	57	HmongDaw	102	Mentawai	141	RomaniSinti
12	ArabicSudaneseSpoken	58	Huambisa	103	MixtecAyutla	143	Sango
15	Avokaya	59	HuavedeSanMateodelMar	104	MofuGudur	145	Sasak
17	AwaPNG	64	IxilNebaj	105	MongolianHalh	146	SauriaPaharia
18	Aymara	65	Javanese	106	Mongondow	150	Shuar
19	AzerbaijaniNorth	68	JurModo	108	Moro	152	Tektiteko
20	Bali	69	Kalabari	109	Mundari	153	TepehuaPisaflores
21	Bekwarra	70	KalmykOirat	110	Munukutuba	154	TepehuaTlachichilco
22	BelizeKriolEnglish	71	Kannada	111	Murle	156	Teribe
24	Bhojpuri	72	Karakalpak	112	Musgu	157	Teso
25	Biate	74	Kashinawa	113	NagaAo	158	Thai
26	BicolanoCentral	75	Kazakh	114	NahuatlGuerrero	159	Themne
28	Bora	76	Kera	115	NahuatlHighlandPuebla	160	Ticuna
29	BruEastern	77	Khasi	116	NahuatlNorthernPuebla	161	Tobelo
30	CakchiquelSouthCentral	79	KicheCunen	117	Napu	162	Tol
31	CandoshiShapra	80	KilivilaKiriwina	118	NigerianPidgin	163	TotonacHighland
34	ChayahuitaShawi	81	KokBorok	120	OromoEastern	166	Tuva
35	ChinantecdeUsila	82	KolamiNorthwestern	121	OtomiMezquital	167	TzeltalBachajon
37	ChujIxtatan	83	Konso	122	Paez	168	Vai
39	DanEast	84	Koorete	123	Palauan	169	Vietnamese
40	Dangaleat	85	KoreanNorth	124	PalaungRuching	170	Waimaha
41	DinkaNortheastern	87	Kui	125	Pamona_BahasaTaa	171	Waiwai
42	Duala	88	Kuman	126	Pampangan	172	WichiLhamtesNoctenWeenhayek
43	Duri	89	Kumyk	127	Parecis	174	Yawa
44	Dutch	90	Kyrgyz	128	Patamona	176	ZapotecSierradeJuarez
45	Eleme	92	Lao	130	Peve		

FIG. 1. Languages corresponding to the numbers.

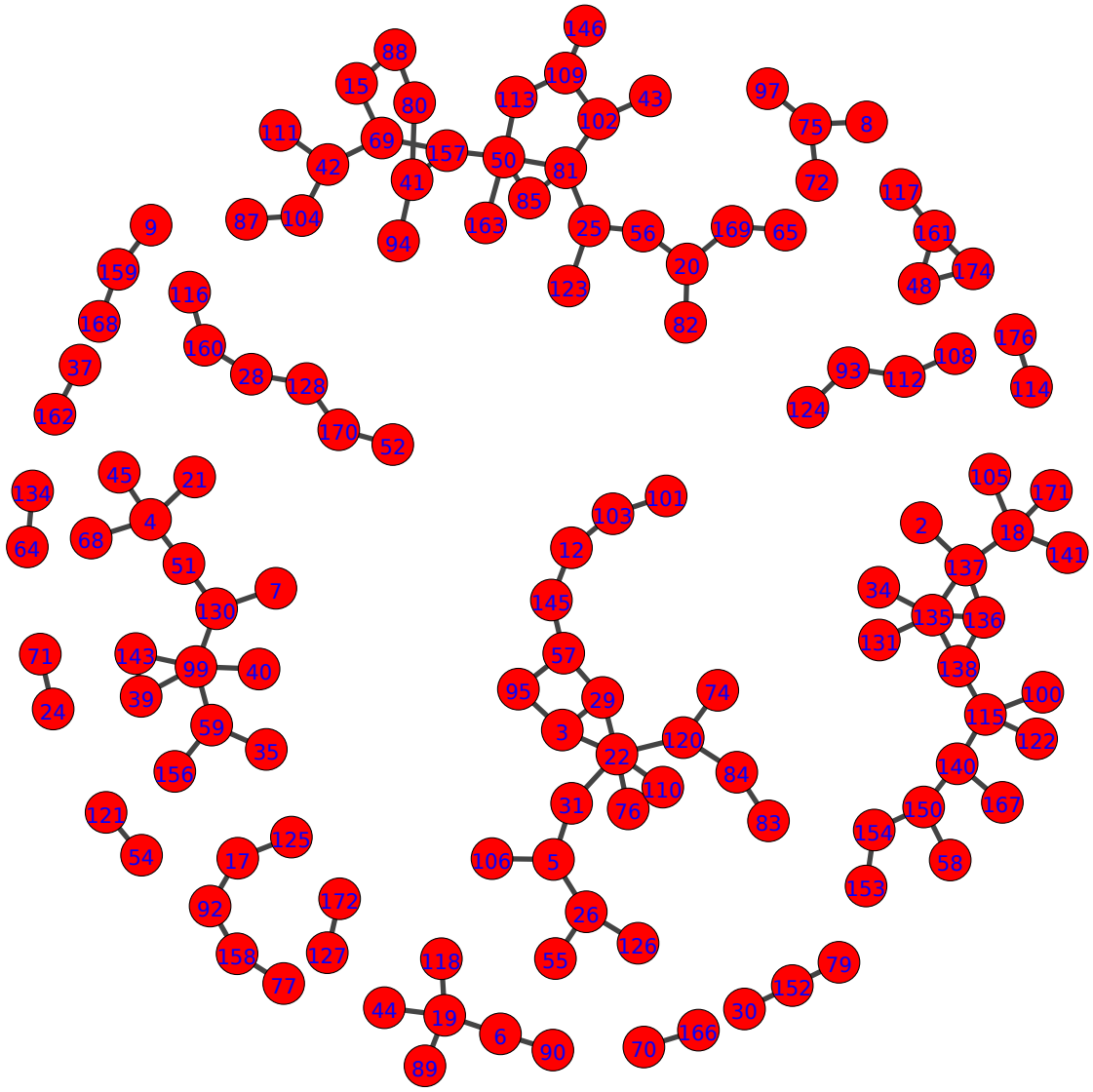


FIG. 2. *Languages connected by positive correlations for the dataset with 176 classes. A small cluster is included in the main text. The complete graph is included in this graphic.*

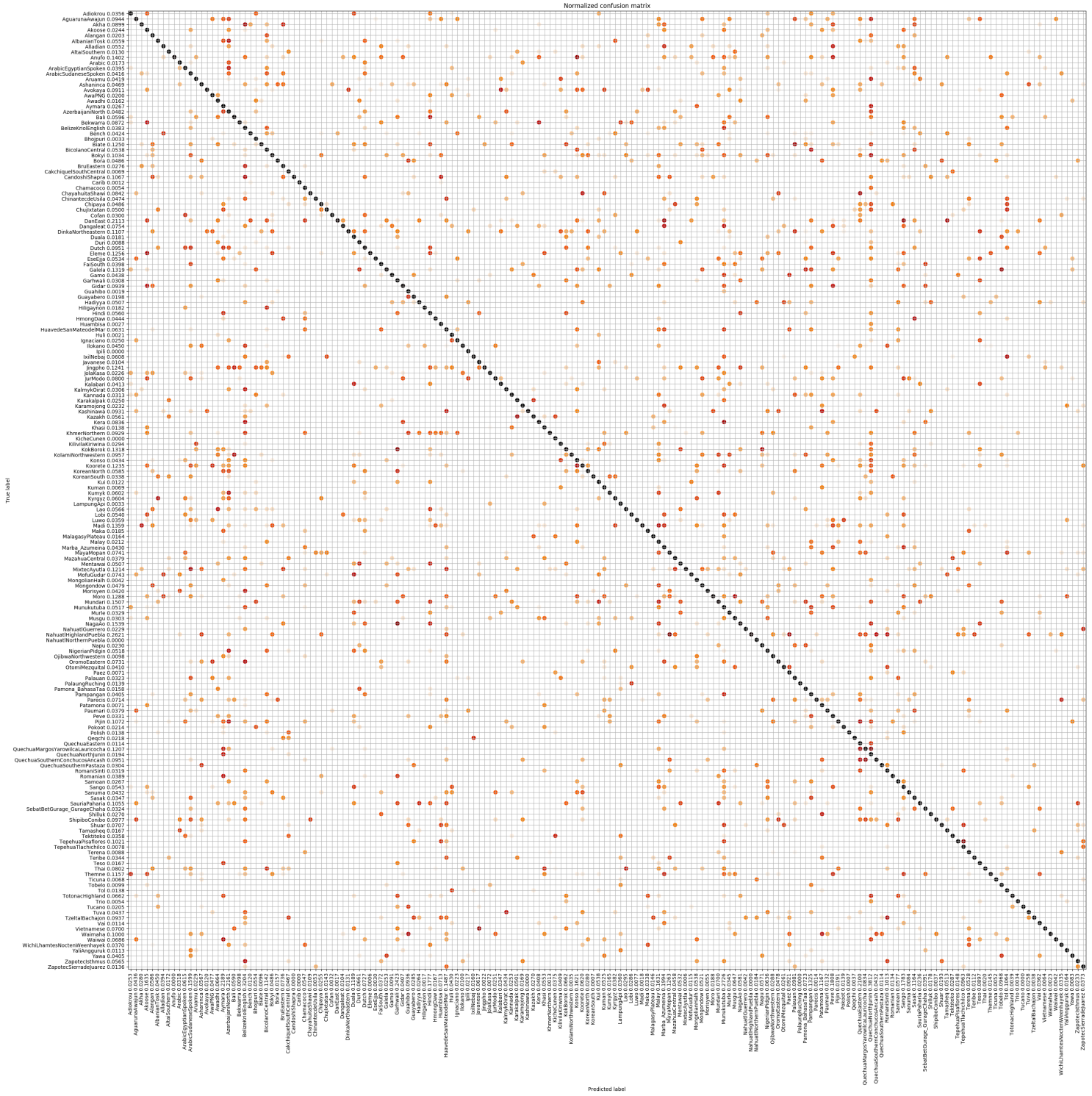
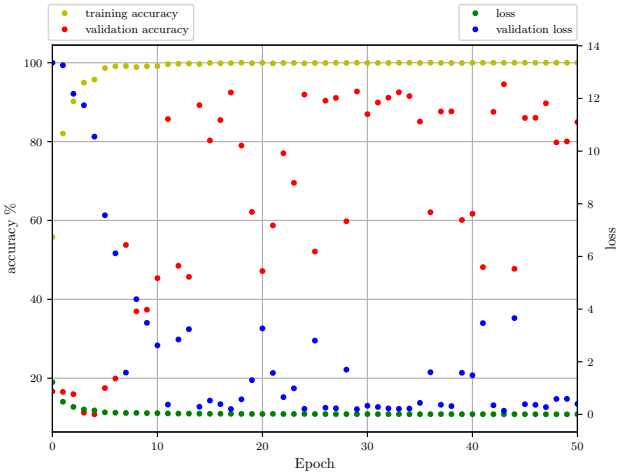
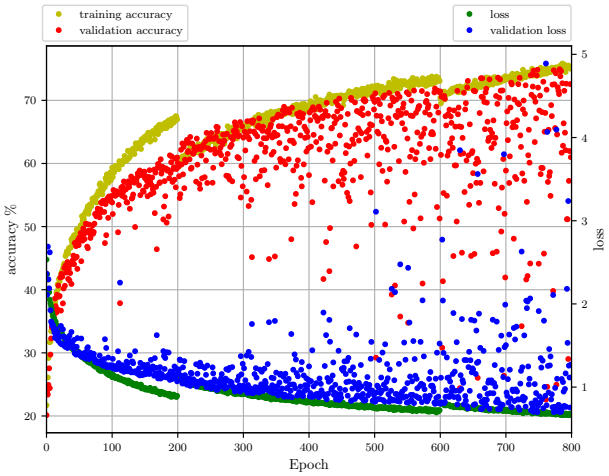


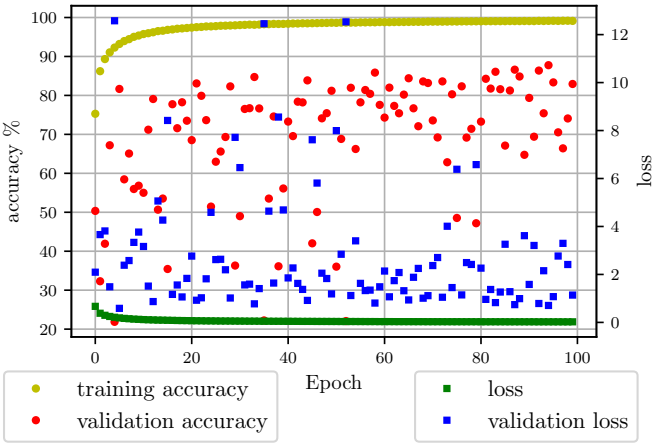
FIG. 3. *Confusion Matrix for the validation of the network trained on the dataset consisting of audio files from the Topcoder challenge with 176 languages. Numbers after the labels give the sum of the corresponding row or column without the diagonal entry. This number gives an indication, how often a language is misclassified as any other one for the rows. For the columns it indicates, how often other languages are wrongfully classified as the language given.*



(a) *topcoder dataset.*



(b) *VoxForge dataset where the dataset was increased at epoch 200 and epoch 600.*



(c) *VoxForge dataset with increased steps per epoch.*

FIG. 4. *Training histories for different datasets and network parameters.*