

Instrument Classification – Report (DeepCNN, Validation Split ~12k)

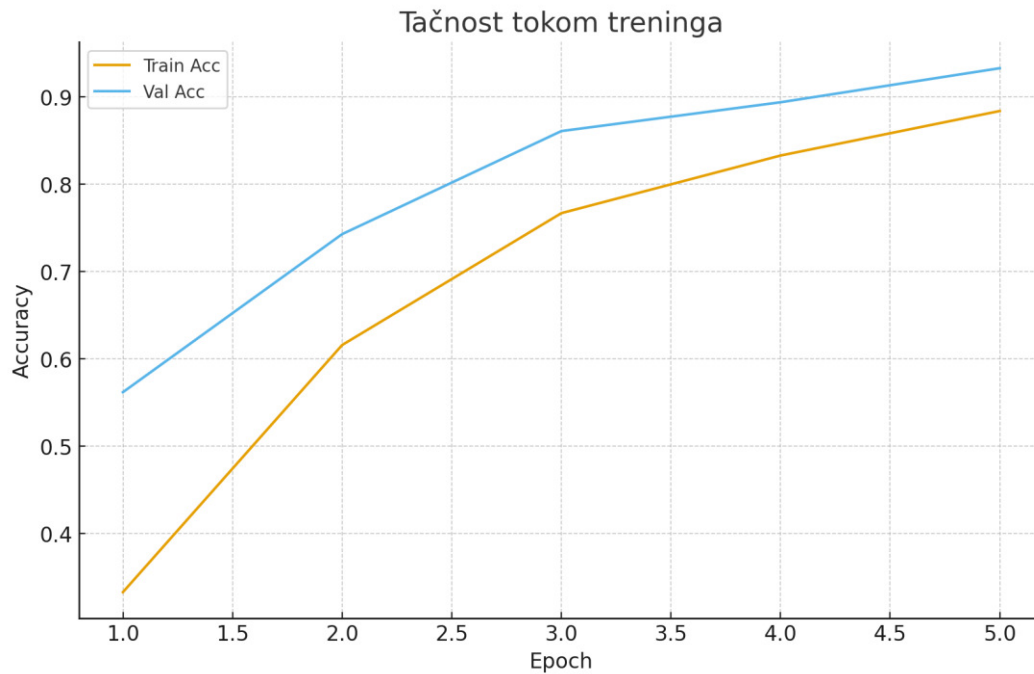
What We Changed

We transitioned from the SmallCNN to a deeper model (DeepCNN): a higher number of channels (32→64→128→256), BatchNorm, Dropout on convolutional blocks, and an additional Fully Connected (FC) layer after Global Average Pooling. The optimizer and loss function remained the same (Adam, CrossEntropy).

Training Trends

Loss and accuracy (5 epochs):





Summary per Epoch:

Ep	Train Loss	Train Acc	Val Loss	Val Acc
1	1.735	0.333	1.286	0.562
2	1.018	0.616	0.859	0.743
3	0.637	0.767	0.539	0.861
4	0.461	0.833	0.365	0.894
5	0.315	0.884	0.270	0.933

Performance per Class (validation)

class	precision	recall	f1	support
bass	0.99	0.82	0.90	160
brass	0.99	0.99	0.99	160
flute	0.99	0.98	0.99	94
guitar	0.73	0.97	0.84	160
keyboard	0.92	0.62	0.76	160

mallet	0.91	0.91	0.91	133
organ	0.97	0.99	0.98	160
reed	0.97	0.98	0.98	144
string	0.95	0.98	0.97	160
vocal	1.00	0.99	0.99	81

Globally: accuracy ≈ 0.93 (N=1412), macro F1 ≈ 0.94 , weighted F1 ≈ 0.94 .

Comments:

There was a significant leap compared to the previous SmallCNN (~ 0.78 F1 $\rightarrow \sim 0.94$ F1). Confusion between related classes (e.g., keyboard \leftrightarrow guitar) still exists, but the overall error rate is now significantly smaller.

Next steps

- Extend training (10 epochs).
- Instead of K=800, use WeightedRandomSampler for class balancing.