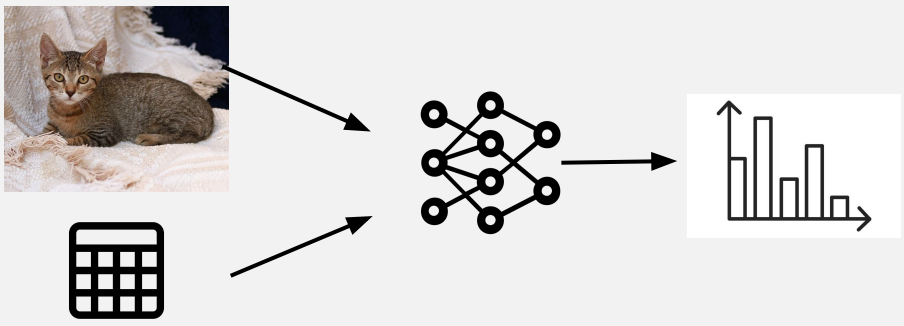
Multimodal Deep Learning for Animal Classification

Tatiana Eliseeva, Mike Krähenbühl, Andrea Staub

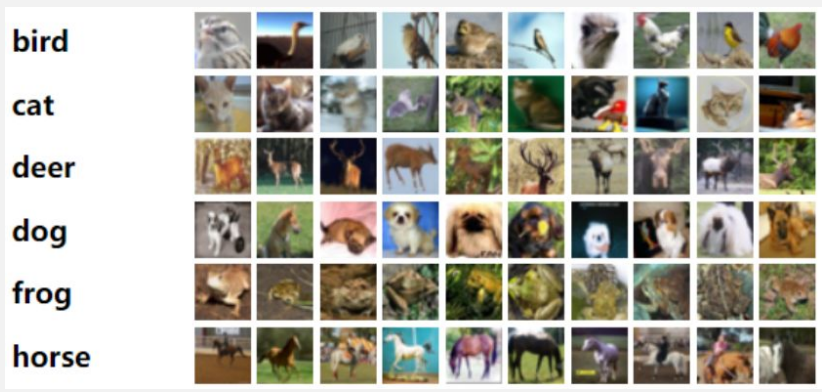
1 Goal

Predicting the probability distribution of animal categories based on image data and tabular data using different NN architectures.



2 Data

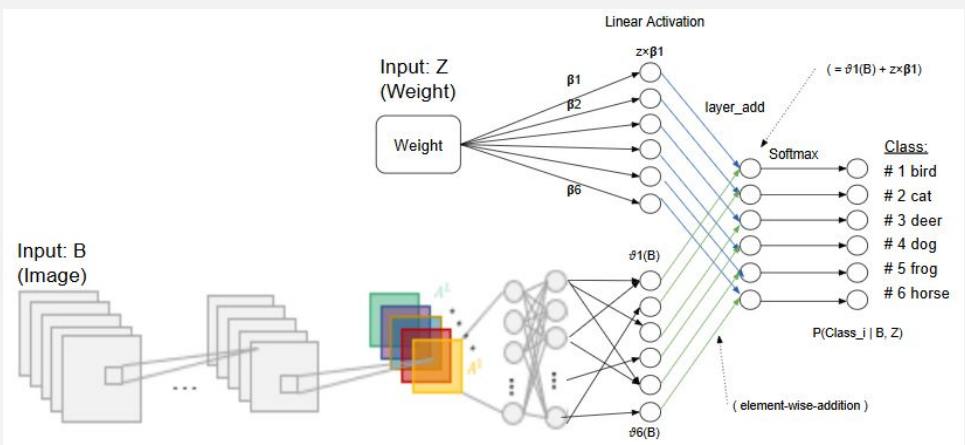
- **Image data:** CIFAR-10 animal images, 32x32 colour images in six classes, with 6’000 images per class.
- **Tabular data:** Randomly generated weight from uniform distribution with reasonable lower and upper bounds for each animal type.



3 Applied Model Architectures

Model Name	Architecture Brief Summary*	Inputs
Model 1 - CNN	“Classic” Image CNN	Images
Model 2 - 1LfcNN-b	Single-Layer fcNN. Contains weights only, softmax activation function without bias	Weights
Model 3 - 1LfcNN	Single-Layer fcNN. Contains weights only, softmax activation function with bias	Weights
Model 4 - Deep-fcNN	Deep fcNN. Contains weights only, ReLu kern with bias	Weights
Model 5 - CNN + 1LfcNN	CNN + Single-Layer fcNN	Images and weights
Model 6 - CNN + Deep-fcNN	CNN + Deep fcNN	Images and weights

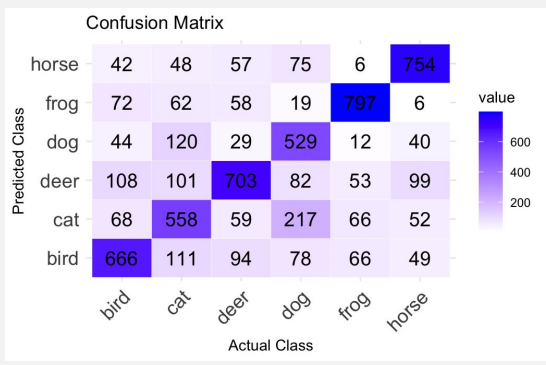
*Always softmax activation at output nodes



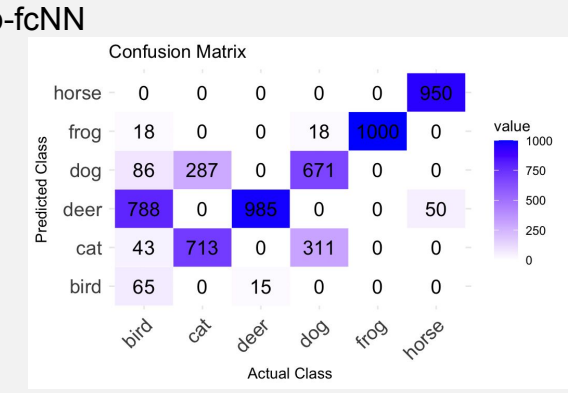
4 Results

Model	Accuracy	Loss
CNN	0.674	0.902
1LfcNN-b	0.167	1.766
1LfcNN	0.446	1.451
Deep-fcNN	0.757	0.537
CNN + 1LfcNN	0.713	0.792
CNN + Deep-fcNN	0.871	0.344

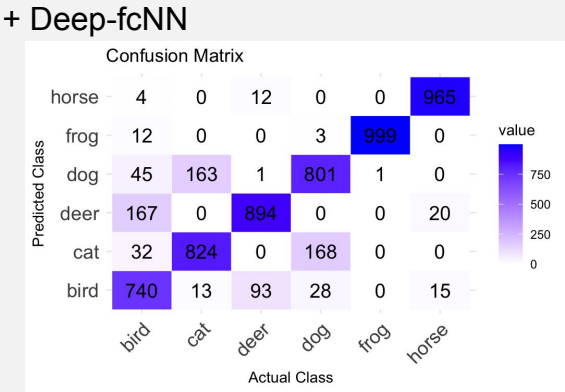
- CNN



- Deep-fcNN



- CNN + Deep-fcNN



5 Conclusion

- Single-Layer NN has too low accuracy
- Deep NN allow to capture the features needed for proper categorisation
- Already one layers of fcNN that allows to process tabular data, can significantly increase the accuracy
- Best results shows combination of CNN and Deep-fcNN