The Motivation:

A neural network is trained to distinguish pictures of cats and dogs provided by the “Kaggle Cats and Dogs Dataset“ from Microsoft (`https://www.microsoft.com/en-us/download/details.aspx?id=54765`).
In a first approach, the data will be transformed into “gray“ images before training, eliminating effectively all colour. After this, the same network structure will be trained on the original “rgb“ coloured image (i.e. red-green-blue). The main idea here is exploring, to what extent the feature of colour will affect the outcome of fitness of the network.

Structure of the Network:

The network consists of two convolutional layers (), () followed by a dense and a ““ layer.

Training and Validation: 80

Both networks will be trained on images of the size of 10x10, 30x30 and 100x100 pixels in the course of one, three and ten epochs. Here, the training data will comprise 80 % of the set i.e. (~20,000 images), consequently validation will be done with ~5,000 images).

Results:

Image Size: 10x10

Epochs	Loss	Acc.	Val.-Loss	Val.-Acc.	“k“	Percentage
1	0.5927	0.7033	0.5411	0.7295	315	1.26
3	0.5437	0.7284	0.5019	0.7527	4656	18.66
10	0.4596	0.7804	0.4354	0.7956	4233	16.97
1	0.5943	0.7034	0.5347	0.7317	315	1.26
3	0.5202	0.7438	0.4926	0.7679	4656	18.66
10	0.4062	0.8150	0.3937	0.8267	4233	16.97

Image Size: 30x30

Epochs	Loss	Acc.	Val.-Loss	Val.-Acc.	“k“	Percentage
1	0.5473	0.7285	0.4667	0.7926	4513	18.09
3	0.4383	0.7985	0.3802	0.8299	3564	14.29
10	0.2916	0.8764	0.2665	0.8878	2687	10.77
1	0.5629	0.7211	0.5228	0.7760	4513	18.09
3	0.4477	0.7943	0.4233	0.8162	3564	14.29
10	0.2707	0.8877	0.2508	0.8992	2687	10.77

Image Size: 100x100

Epochs	Loss	Acc.	Val.-Loss	Val.-Acc.	“k“	Percentage
1	0.5748	0.7200	0.4435	0.7988	3675	14.73
3	0.4077	0.8165	0.3464	0.8511	3165	12.69
10	0.1138	0.9565	0.1084	0.9689	1052	4.21
1	0.5637	0.7311	0.4995	0.7703	3675	14.73
3	0.3790	0.8376	0.3267	0.8697	3165	12.69
10	0.1569	0.9376	0.1436	0.9453	1052	4.21

The collected benchmarks are displayed in “table1“. The letter “k“ stands for the number of images of the whole data set where both models differ in their predictions, “percentage“ is just the relative part of k with respect to the whole dataset. The upper, resp. lower three rows in each table represent the network trained on “gray“, resp. coloured images.
For each image size the loss in both settings naturally declines with increasing number of epochs trained. Here, the larger the image size, the steeper the decrease. Interestingly, the loss after ten epochs of the “rgb“-model is lower than the loss of the “gray“-model for the 10x10 and 30x30 but not for 100x100 pixels. Indeed, the “rgb“-model also outperforms the “gray“-model in terms of accuracy only in the 10x10 and 30x30 case after 10 epochs of training.
When training on any image size for only one epoch, both models perform almost equally. In the case of three epochs the “rgb“-model outperforms the “gray“-model again for 10x10- and 30x30- but will not for the 100x100-images.

In terms of the number of images (“k“) for which the predictions of the two models differ, one can see that for the smallest image size (10x10) and

after only one epoch of training both are very according in their decisions (1.26 %), i.e. both are effectively making the same mistakes. The percentage then increases rapidly when trained for more epochs.
For the 30x30 image size, they differ the most after only one epoch of training which then decreases after three and ten epochs, meaning they will tend to rather make the same mistakes, the more epochs trained.
The same trend but more steep is to recognise in the last table for the 100x100 size, where after 10 epochs the percentage is really low (4.21 %).

Hypothesis:

One thing one may take from this is the following: for low-dimensional data (10x10 and 30x30 pixels) the model trained on coloured images performs better than the model trained on “gray“ images given sufficiently many epochs of training. But when the images become more detailed, colour seems rather to distract the network. Mathematically, one may interpret accuracy as a function $P : \mathbb{N}^2 \to [0, 1]$ depending on image size $s \in \mathbb{N}$ and epochs $e \in \mathbb{N}$, i.e. $P(s, e)$. Let furthermore $P_{gray}(s, e)$ and $P_{rgb}(s, e)$ the respective performance functions of the “gray“- and “rgb“-model. Then there is a $s_0 \in \mathbb{N}$ such that for

$$D(s, e) := P_{gray}(s, e) - P_{rgb}(s, e)$$

it holds:

- $D(s, e) \leq 0$ for $s \leq s_0, e >> 0$
- $D(s, e) \geq 0$ for $s \geq s_0, e >> 0$

A similar representation for loss as a function should also exist.