The Motivation:

A convolutional neural network is trained to distinguish pictures of cats and dogs provided by the “Kaggle Cats and Dogs Dataset“ from Microsoft (approx. 25,000 images; URL: `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 motivation here is exploring to what extent the feature of colour will affect the fitness of the network.

Structure of the Network:

The models consist both of two sequences of Convolutional layers of dimension 64 with strides (3,3), activation function ReLU and pooling size (2,2) followed by a Dropout layer with rate 0.15.
Afterwards a Flatten layer as well as two Dense layers of dimension 64 and 1, respectively; the activation function is given by the Sigmoid function.
Both models will be compiled using the “binary crossentropy“-loss, the “adam“-optimizer and the “accuracy“-metric.

Training and Validation:

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 in batches of 32 with relative validation size of 20% (~5,000 images). Consequently, the training data will comprise 80% of the set (~20,000 images).

Results:

Image Size: 10x10

Epochs	Loss	Acc.	Val.-Loss	Val.-Acc.	“k“	Percentage
1	0.5927	0.7033	0.5411	0.7295	$\boxed{315}$	$\boxed{1.26}$
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	$\boxed{315}$	$\boxed{1.26}$
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	$\boxed{0.1084}$	$\boxed{0.9689}$	$\boxed{1052}$	$\boxed{4.21}$
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	$\boxed{0.1436}$	$\boxed{0.9453}$	$\boxed{1052}$	$\boxed{4.21}$

The collected data is displayed above. The letter “k“ stands for the number of images (with respect to the whole data set) where both models differ in their predictions, “percentage“ is just the relative part of k. For each test setup (comp. the tables) two images (in colour) have been picked for which both models made different decisions (the respective predictions are written on top of each image). The upper, resp. lower three rows in each table represent the model trained on “gray“, resp. “rgb“ images.

For each image size the loss of both networks naturally declines with increasing number of epochs trained. Here, the larger the image size, the steeper the decrease. 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 regardless of the image size. Interestingly, the validation loss after ten epochs of the “rgb“-model is lower than the loss of the “gray“-model for size 10x10 and 30x30, but not for size 100x100. Furthermore, the “rgb“-model outperforms the “gray“-model in terms of validation accuracy only in the 10x10 and 30x30 case after 10 epochs of training, not in the 100x100 case.

When it comes to 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, meaning the models are making many more different mistakes, although the amount of errors in total nearly stays the same (judging by the validation accuracy).
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 accord in their predictions (right or wrong), 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 may take the following from this: for low-dimensional data (10x10 and 30x30 pixels) the model trained on “rgb“-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 in its predictions, supporting the idea that it is a less important feature. When it comes to differentiating pictures of cats and dogs of reasonable size, other features may be more characteristic than colour.
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. $P(s, e)$. If $P_{gray}(s, e)$ and $P_{rgb}(s, e)$ are the respective accuracy 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 $s \leq s_0, e \gg 0$
- $D(s, e) \geq 0$ for $s \geq s_0, e \gg 0$

A similar representation for the loss function should exist as well.