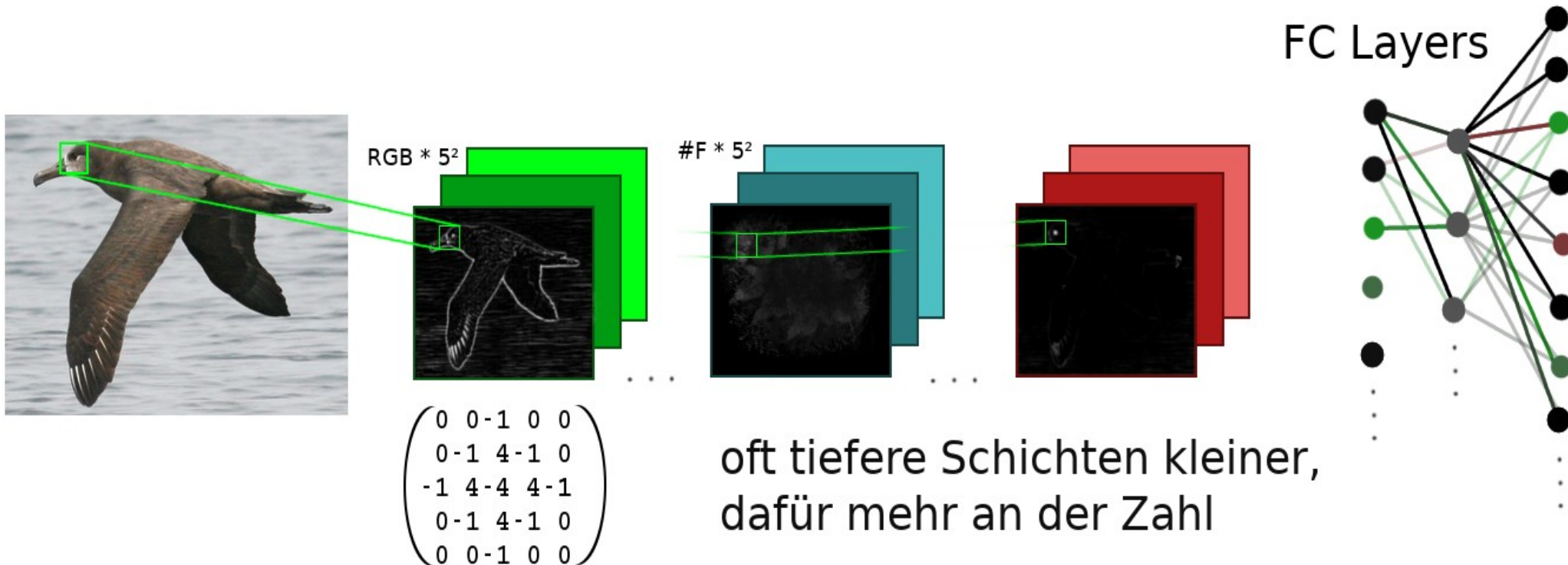


Learning Multi-Attention CNNs for Fine Grained Image Recognition

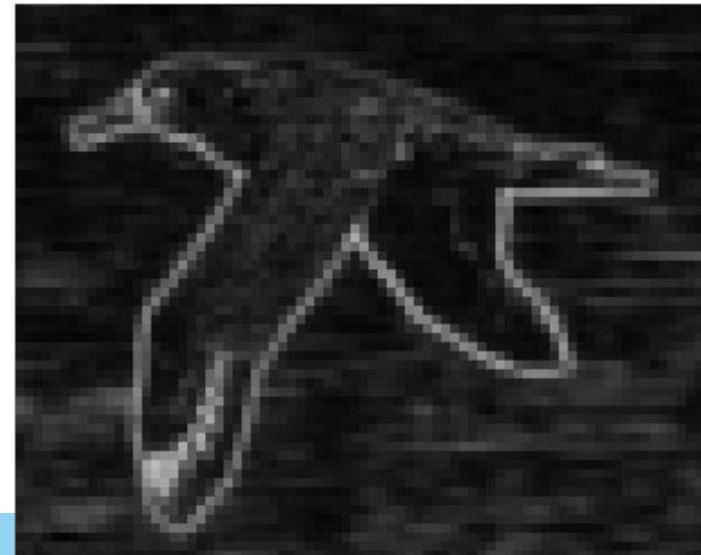


Convolutional Neural Networks



Pooling

- Oft Maximum der Pixel-Nachbarschaft
- Zur Vergrößerung der Nachbarschaft der Features, für bedeutendere Features
- Ohne mehr Ebenen weniger Inputs für die FC-Layer



Deep Dream

- Verstärkung von Features durch Backpropagation
- Ermöglicht abgewandelt auch Stilübertragungen



Selbst ausprobieren: deepdreamgenerator.com

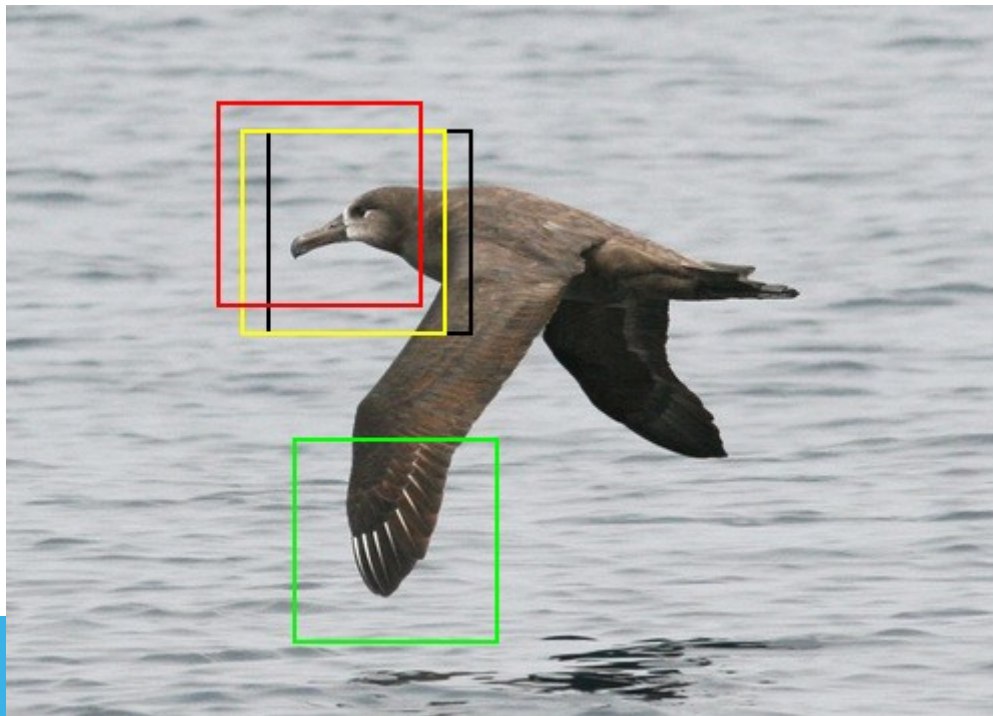
Klassifikation

- I.d.R. entscheidet der stärkste Ausschlag
- Einzelner FC-Layer ist wie Korrelationstabelle
- Zum Trainieren (Optimieren) Fehlerfunktion erforderlich: Softmax gerne verwendet, täuscht allerdings Sicherheit vor:

$$Score_k \leftarrow \frac{\exp(Score_k)}{\sum_i \exp(Score_i)}$$

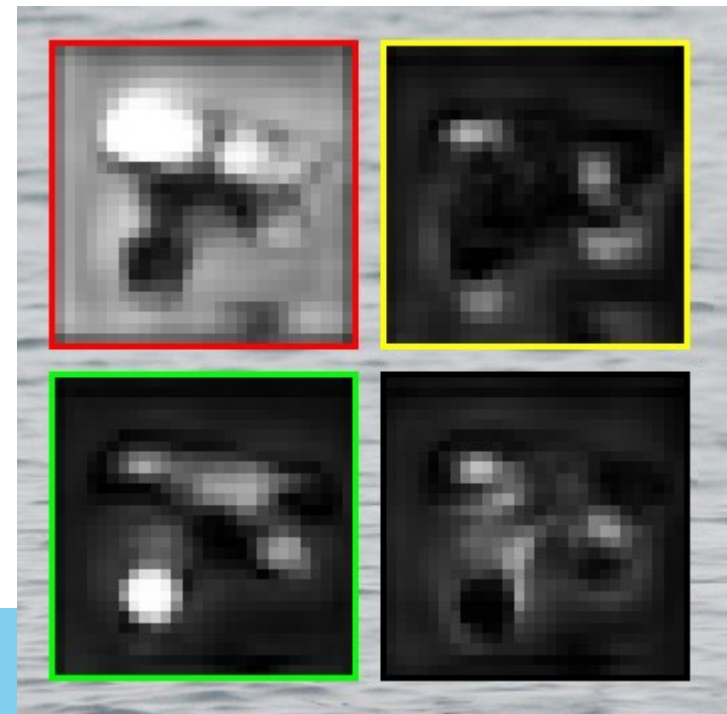
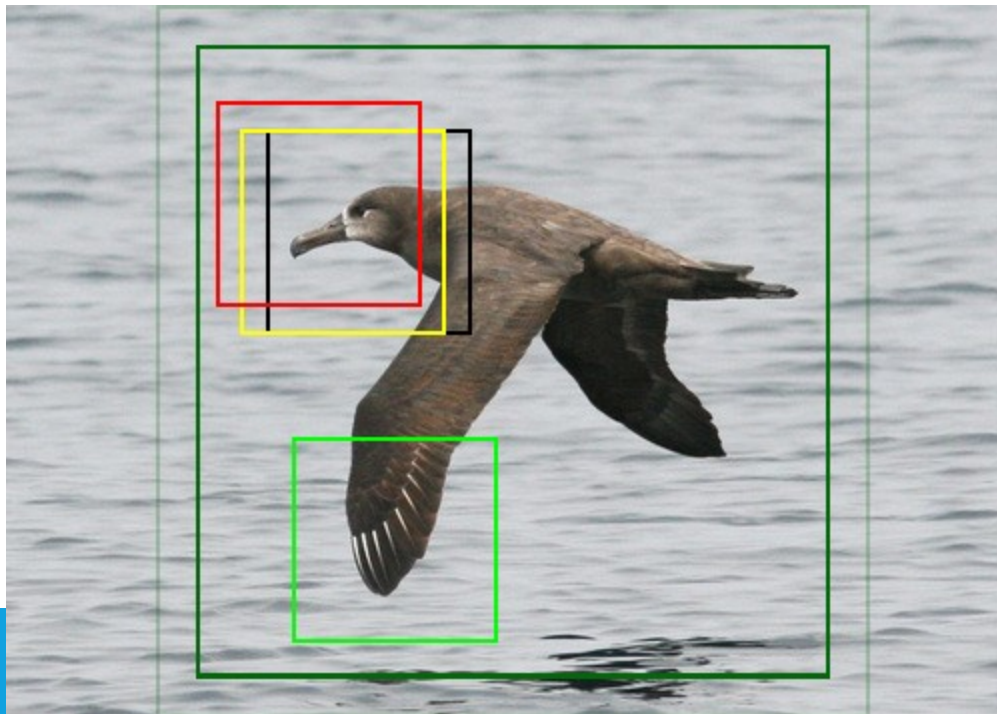
Multi-Attention CNNs

- Eingabe des Originalbildes plus zuzüglich mehrerer Aufmerksamkeitssspots
- Soll Klassifikation erleichtern

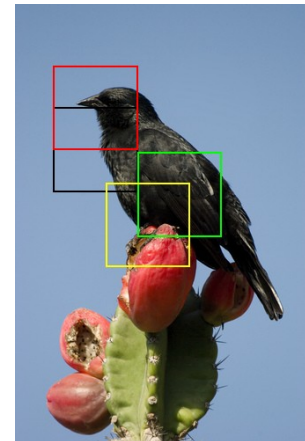
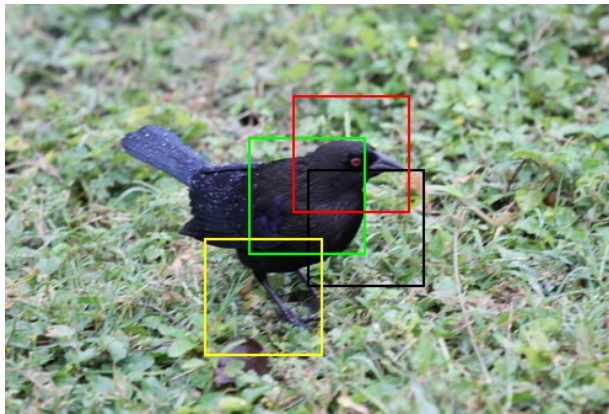
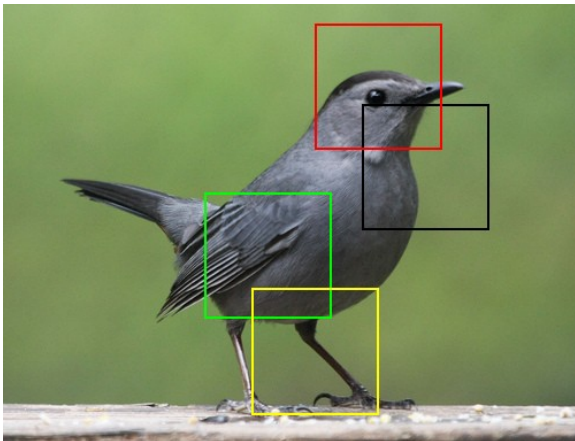
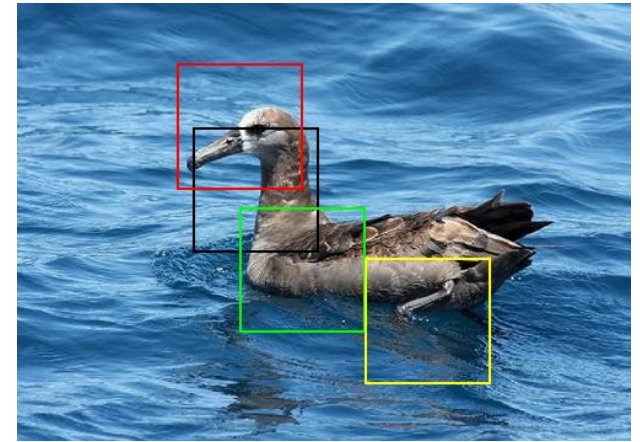
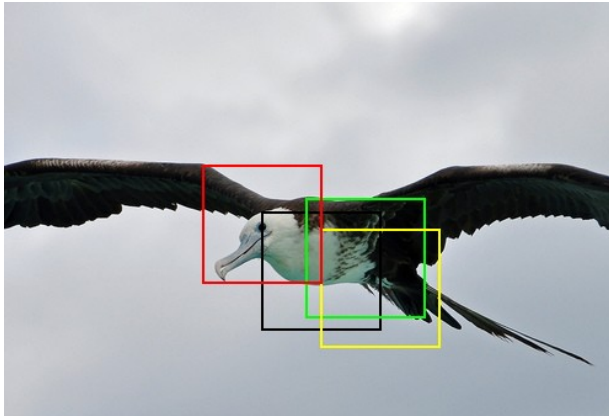
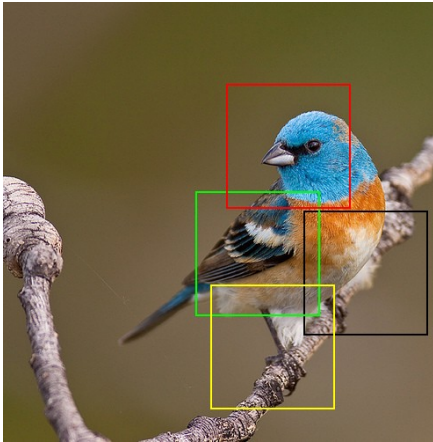


Bestimmung der Spots

- Bisher oft durch Menschen annotiert
- Im Paper nun Aufgabe des Netzwerkes, die Spots auszuwählen: über Heatmaps



Weitere Beispiele



Lernen der Gewichte

- Zwei Möglichkeiten, die sich jedoch ähneln:
 1. Abwechselnd Aufmerksamkeits-Netzwerk und Klassifikator, oder
 2. End-to-end, d.h. im Ganzen
- Fehlerfunktion entscheidend, da sie Lernaufgabe vorgibt:

Was sind gute Ausschnitte?
- Initialisierung von beiden Teilnetzwerken per vortrainiertem ImageNet (VGG-19)

Fehlerfunktion

- Klassifikator-Fehler über Softmax
- Aufmerksamkeitsnetzwerk möglichst:
 1. wenig Überlappung
 2. lokal, d.h. nicht zu breit
- Realisierung durch Linearkombination von:

1.
$$L_{\text{div}} = \sum_{j, j \neq i} \sum_{x, y} A_i(x, y) \cdot A_j(x, y)$$

2.
$$L_{\text{dst}} = \sum_{x, y} A_i(x, y) \cdot ((x - x_c)^2 + (y - y_c)^2)$$

Netzwerk-Zusammenfassung

- Brauchen wir das?...

Ergebnisse

Approach	Train Anno.	Accuracy
PN-CNN(AlexNet) [1]	✓	75.7
Part-RCNN(AlexNet) [34]	✓	76.4
PA-CNN [14]	✓	82.8
MG-CNN [27]	✓	83.0
FCAN [18]	✓	84.3
B-CNN (250k-dims) [17]	✓	85.1
Mask-CNN [29]	✓	85.4
TLAN(AlexNet) [31]		77.9
MG-CNN [27]		81.7
FCAN [18]		82.0
B-CNN (250k-dims) [17]		84.1
ST-CNN (Inception net) [10]		84.1
PDFR [35]		84.5
RA-CNN [5]		85.3
MA-CNN (2 parts + object)		85.4
MA-CNN (4 parts + object)		86.5

Quellen

Literatur:

- Das Paper selbst
- Verschiedene allgemeine Seiten zu z.B. CNNs wie towardsdatascience.com, Wikipedia

Bilder:

- Als Vorlage
<https://i0.wp.com/vinodsblog.com/wp-content/uploads/2018/10/CNN-2.png>
- Generiert mit deepdreamgenerator.com