

# Hinter den Kulissen visueller KI

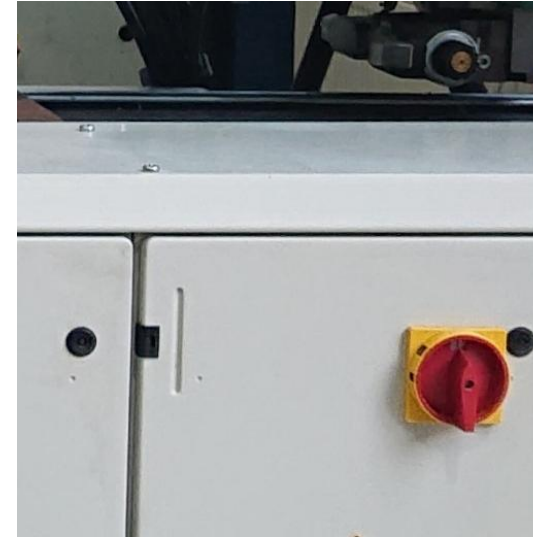
Grundbausteine aktueller KI Modelle in der Bilderkennung

Oliver Kempf M. Sc.



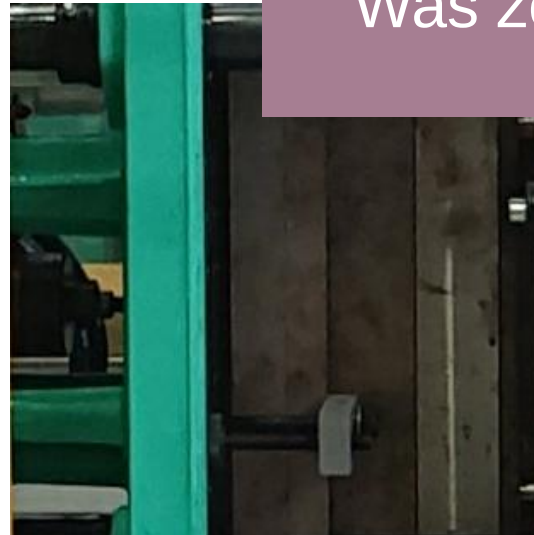
Kofinanziert von der  
Europäischen Union

# Bildverständnis



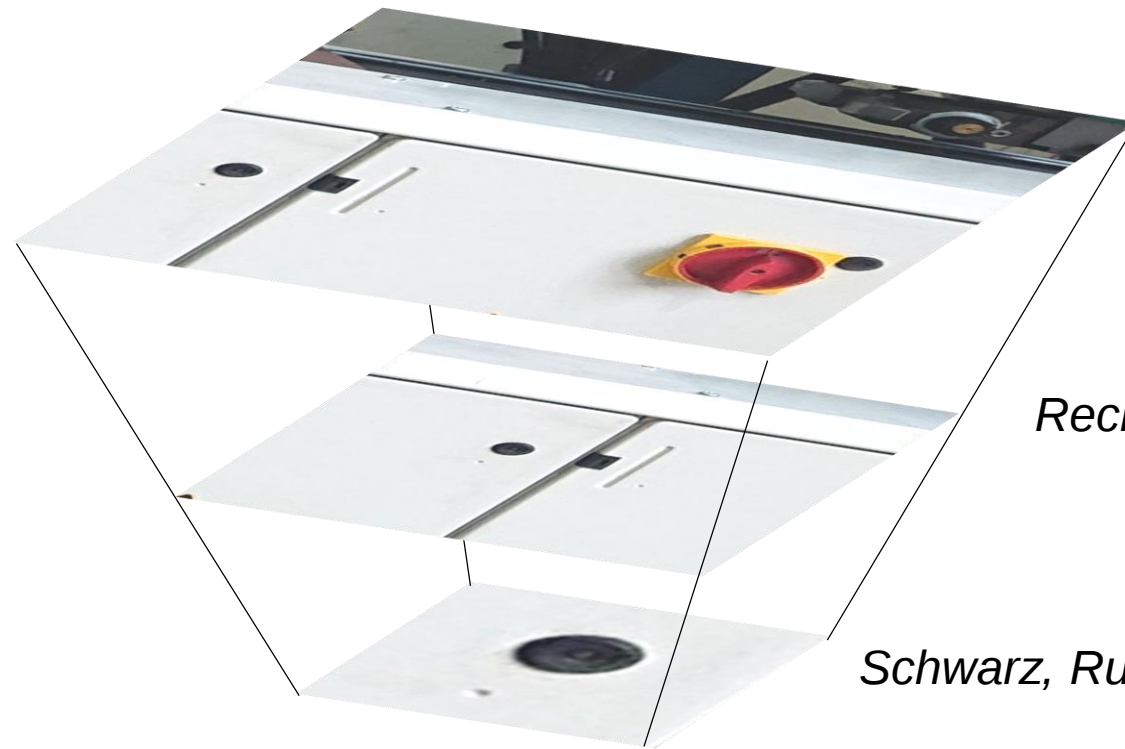


Was zeigen die Bildausschnitte?





# Semantische Hierarchie des Sehens



*Objekt:  
Hauptschalter*

*Rechtwinklige Anordnung*

*Schwarz, Rund, Homogene Fläche*

Interpretation

Industrieanlage?

Schaltschrank?

Mehrteiliges Objekt,  
Technisch

Geometrie und  
Hintergrund

# Robustes Sehen



Original



Schärfe



Perspektivische Verzerrung



Verdeckung



Größe



Helligkeit & Kontrast

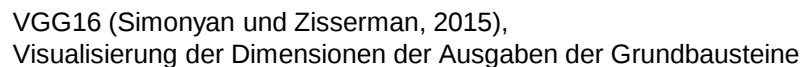


Rotation



DUETT

- 



The diagram illustrates the U-Net architecture for object detection, showing the flow of data from input to output, with dimensions and operations at each stage.

**Input:** Image (640x640x3)

**Encoder (Left Side):**

- Conv 0:**  $k=3, s=2, p=1$  (P1) → 320x320x64x
- Conv 1:**  $k=3, s=2, p=1$  (P2) → 160x160x128x
- C2f 2:**  $\text{shortcut}=\text{True}, n=3 \times d$  → 160x160x128x
- Conv 3:**  $k=3, s=2, p=1$  (P3) → 80x80x256x
- C2f 4:**  $\text{shortcut}=\text{True}, n=3 \times d$  → 80x80x256x (Stride=8)
- Conv 5:**  $k=3, s=2, p=1$  (P4) → 40x40x512x
- C2f 6:**  $\text{shortcut}=\text{True}, n=3 \times d$  → 40x40x512x (Stride=16)
- Conv 7:**  $k=3, s=2, p=1$  (P5) → 20x20x512x
- C2f 8:**  $\text{shortcut}=\text{True}, n=3 \times d$  → 20x20x512x
- SPPF 9:** → 20x20x512x (Stride=32)

**Decoder (Right Side):**

- Upsample 10:** 20x20x512x → 40x40x512x
- C2f 11:**  $\text{shortcut}=\text{False}, n=3 \times d$  → 40x40x512x
- Concat 12:** 40x40x512x (from C2f 11) + 40x40x512x (from Conv 5) → 40x40x512x
- Upsample 13:** 40x40x512x → 80x80x512x
- C2f 14:**  $\text{shortcut}=\text{False}, n=3 \times d$  → 80x80x512x
- Concat 15:** 80x80x512x (from C2f 14) + 80x80x256x (from Conv 3) → 80x80x512x
- Upsample 16:** 80x80x512x → 160x160x512x
- C2f 17:**  $\text{shortcut}=\text{False}, n=3 \times d$  → 160x160x512x
- Concat 18:** 160x160x512x (from C2f 17) + 160x160x128x (from Conv 1) → 160x160x512x
- Upsample 19:** 160x160x512x → 320x320x512x
- C2f 20:**  $\text{shortcut}=\text{False}, n=3 \times d$  → 320x320x512x
- Concat 21:** 320x320x512x (from C2f 20) + 320x320x64x (from Conv 0) → 320x320x512x

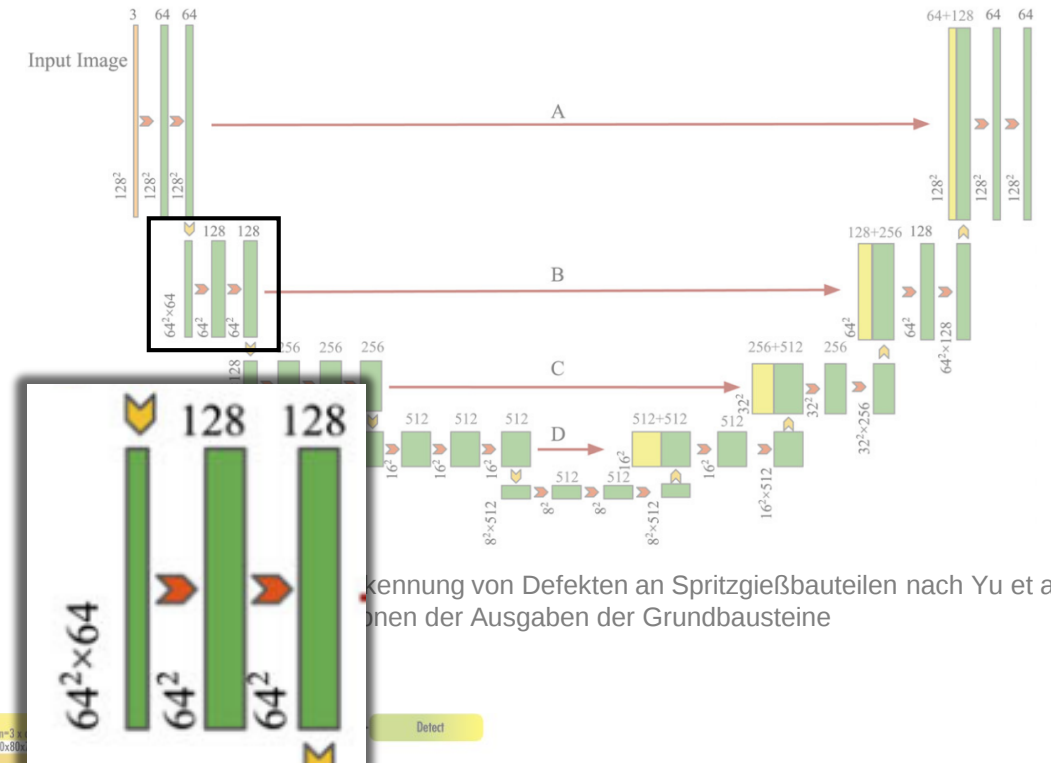
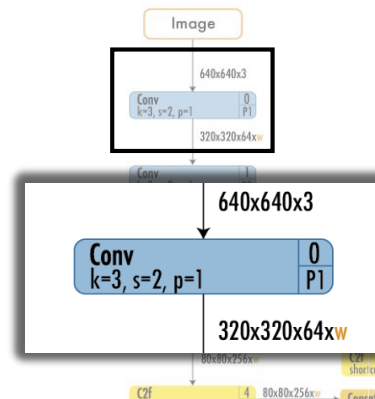
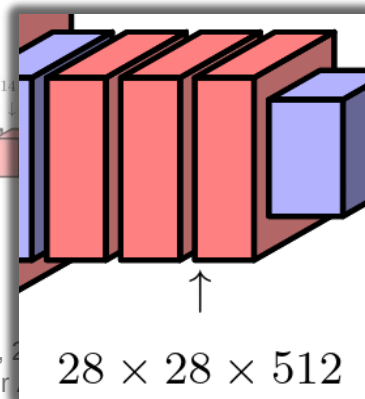
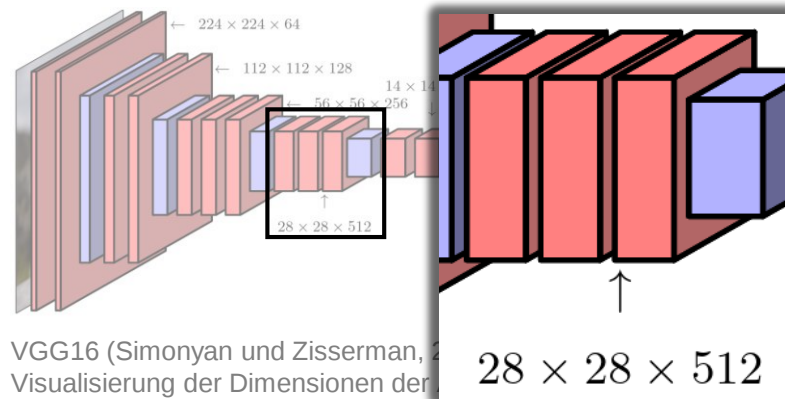
**Output:** Detect (320x320x512x)

**Note:** height x width x channel

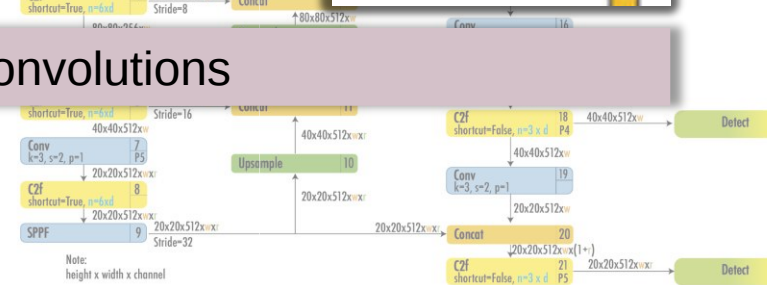
 Kofinanziert von der Europäischen Union

# Grundbaustein: Convolutions

- Historisch: CNNs (o. ConvNets) eigene Architektur
- Hier: CNNs = alle Architekturen, die auf den Grundbaustein *Convolution* zurückgreifen



## Convolutions

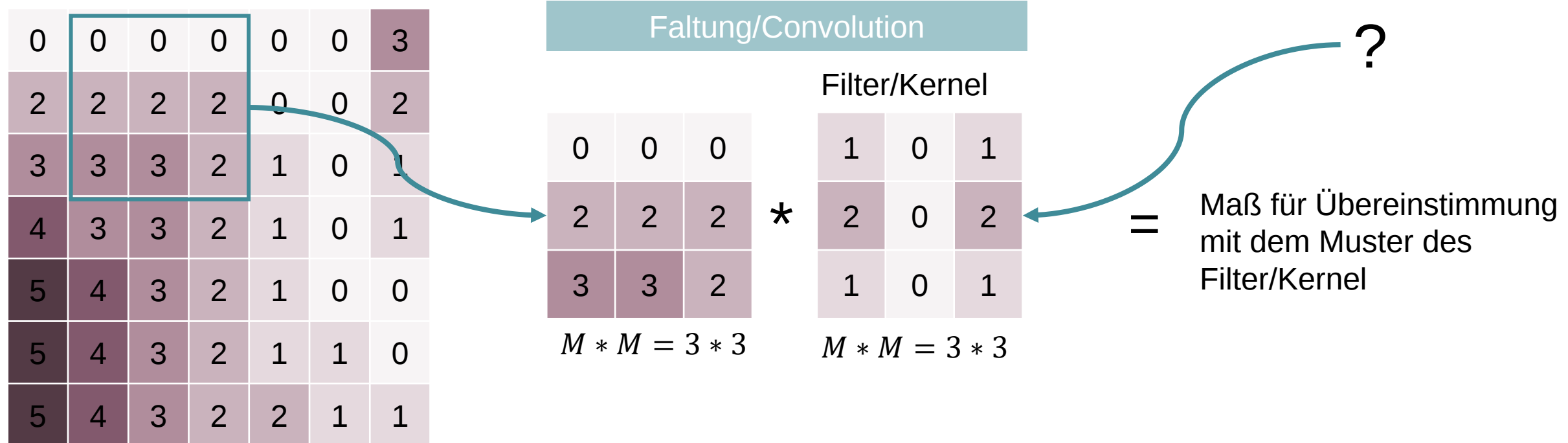


Erkennung von Defekten an Spritzgießbauteilen nach Yu et al. (2024),  
Anwendung der Ausgaben der Grundbausteine

# Convolutional Neural Networks (CNN)

- Netzwerke auf Basis von Convolutions, dt. Faltung
- Diskrete, zweidimensionale Faltung
- Erreicht maximalen Wert, wenn Eingabe und Filter übereinstimmen

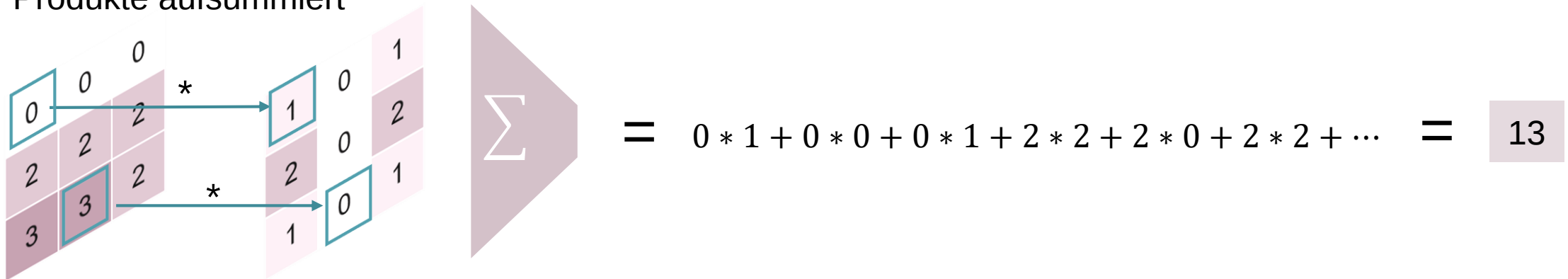
Bildausschnitt



# CNN: Convolutions

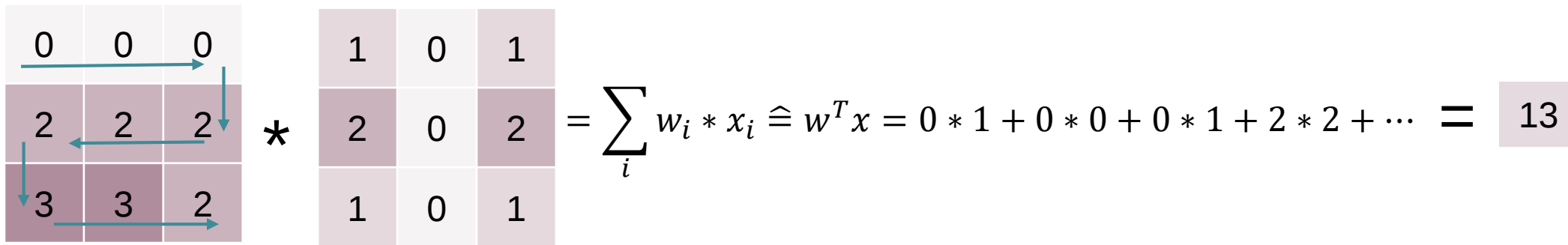
- Netzwerke auf Basis von Convolutions, dt. Faltung
- Diskrete, zweidimensionale Faltung
- Erreicht maximalen Wert, wenn Eingabe und Filter übereinstimmen

Elementweise multipliziert,  
Produkte aufsummiert



# CNN: Convolutions

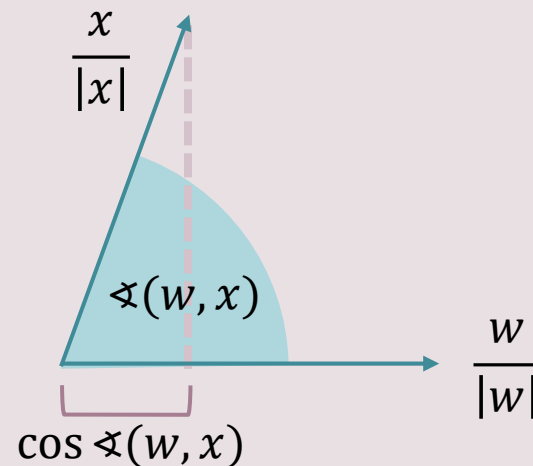
- Netzwerke auf Basis von Convolutions, dt. Faltung
- Diskrete, zweidimensionale Faltung
- Erreicht maximalen Wert, wenn Eingabe und Filter übereinstimmen


$$\begin{bmatrix} 0 & 0 & 0 \\ 2 & 2 & 2 \\ 3 & 3 & 2 \end{bmatrix} * \begin{bmatrix} 1 & 0 & 1 \\ 2 & 0 & 2 \\ 1 & 0 & 1 \end{bmatrix} = \sum_i w_i * x_i \hat{=} w^T x = 0 * 1 + 0 * 0 + 0 * 1 + 2 * 2 + \dots = 13$$

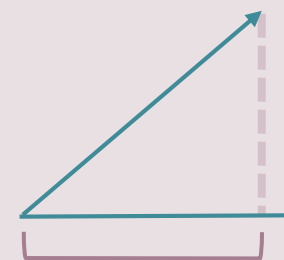
## Exkurs

Wo wird das  
Skalarprodukt maximal?

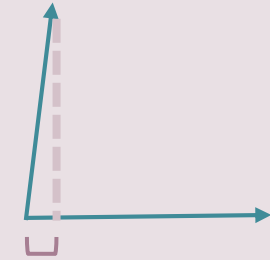
$$w^T x = \cos \angle(w, x) * |w| * |x|$$



Ähnliche Richtung



Abweichende Richtung



# CNN: Räumliche Zuordnung

Bildausschnitt

|   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 3 |
| 2 | 2 | 2 | 2 | 0 | 0 | 2 |
| 3 | 3 | 3 | 2 | 1 | 0 | 1 |
| 4 | 3 | 3 | 2 | 1 | 0 | 1 |
| 5 | 4 | 3 | 2 | 1 | 0 | 0 |
| 5 | 4 | 3 | 2 | 1 | 1 | 0 |
| 5 | 4 | 3 | 2 | 2 | 1 | 1 |

|   |   |   |
|---|---|---|
| 0 | 0 | 0 |
| 2 | 2 | 2 |
| 3 | 3 | 2 |

\*

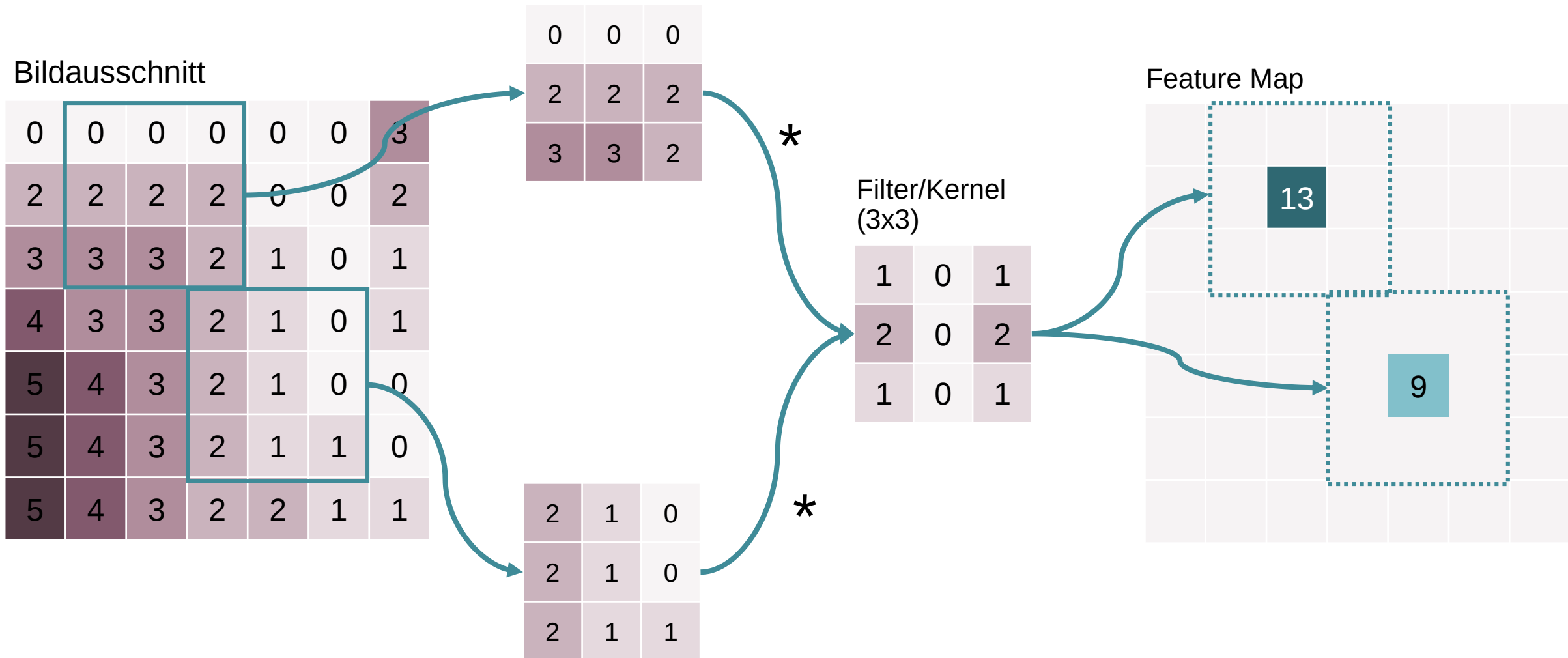
Filter/Kernel  
(3x3)

|   |   |   |
|---|---|---|
| 1 | 0 | 1 |
| 2 | 0 | 2 |
| 1 | 0 | 1 |

Feature Map, dt. Merkmals-Karte

|  |  |    |  |  |  |  |
|--|--|----|--|--|--|--|
|  |  |    |  |  |  |  |
|  |  |    |  |  |  |  |
|  |  | 13 |  |  |  |  |
|  |  |    |  |  |  |  |
|  |  |    |  |  |  |  |
|  |  |    |  |  |  |  |
|  |  |    |  |  |  |  |

# CNN: Beibehaltung räumlicher Verhältnisse



# CNN: Interpretation der Feature Map

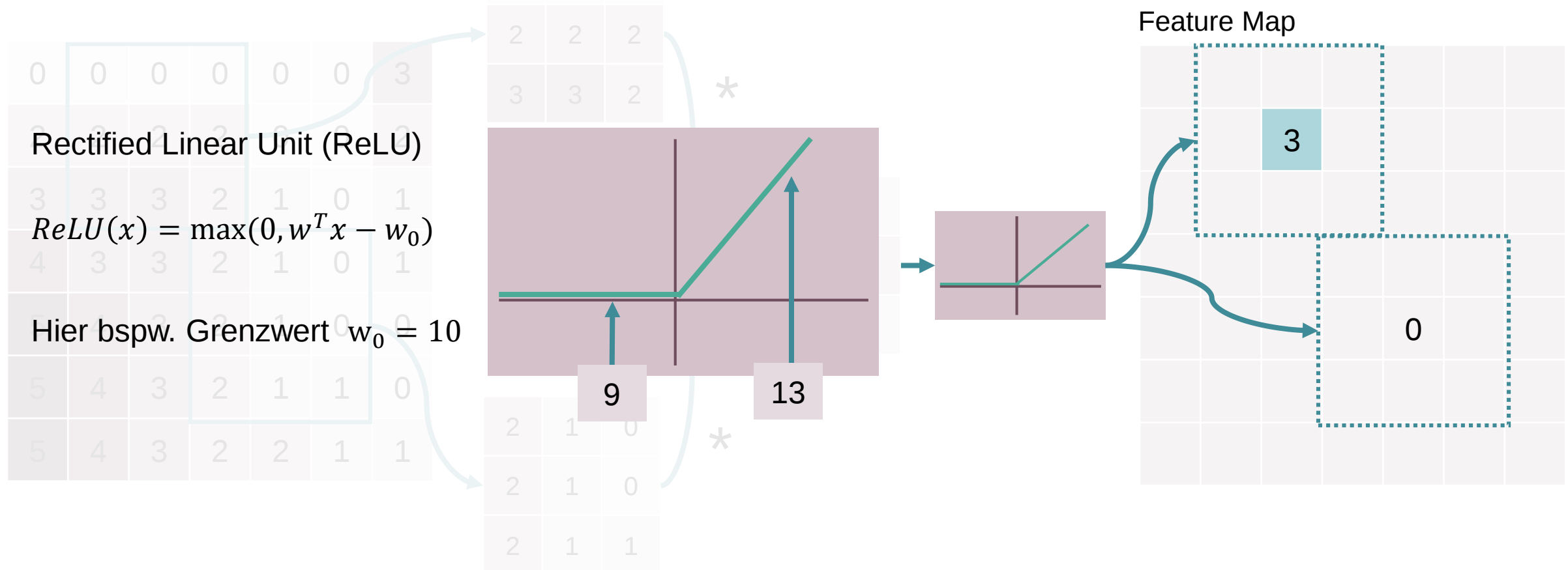
Filter für horizontale Kanten, Filter =  $[101, 202, 101]$



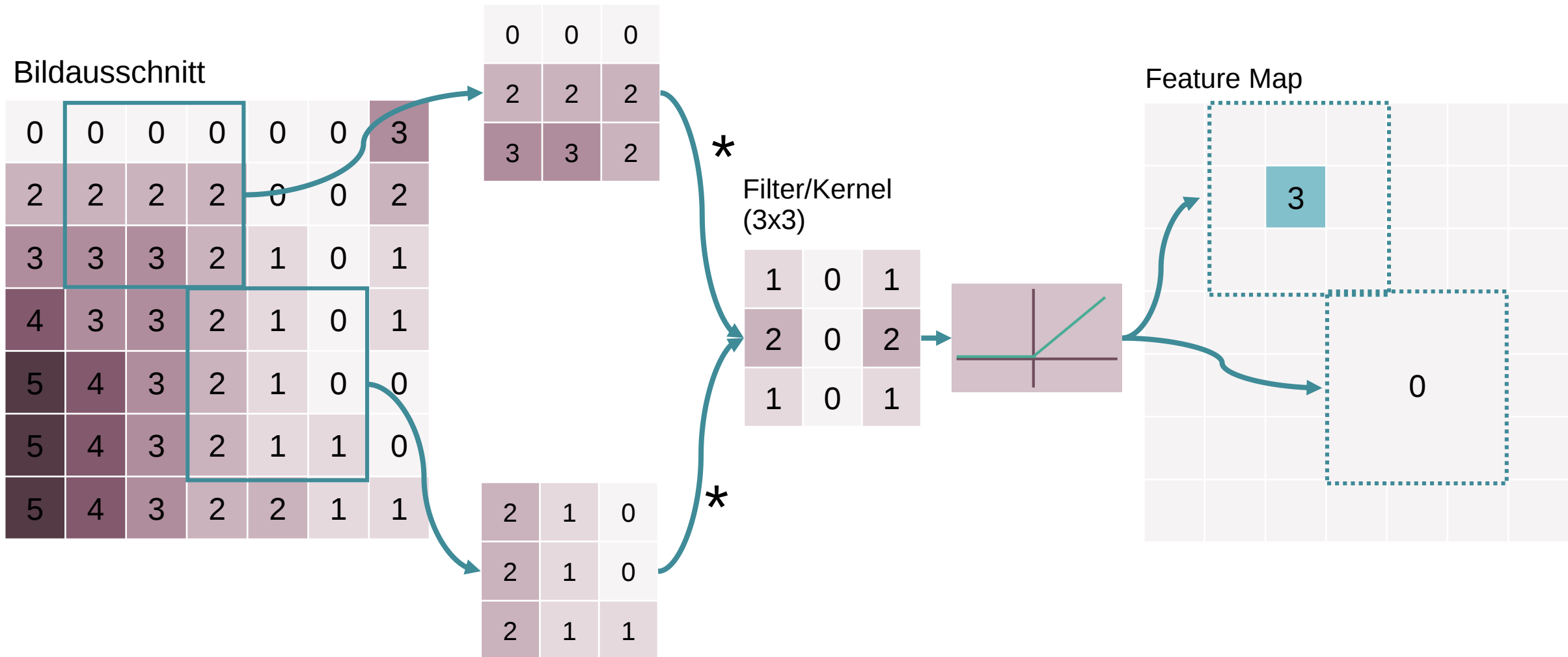
Welches Feature ist rechts dargestellt?

# CNN: Aktivierungsfunktion

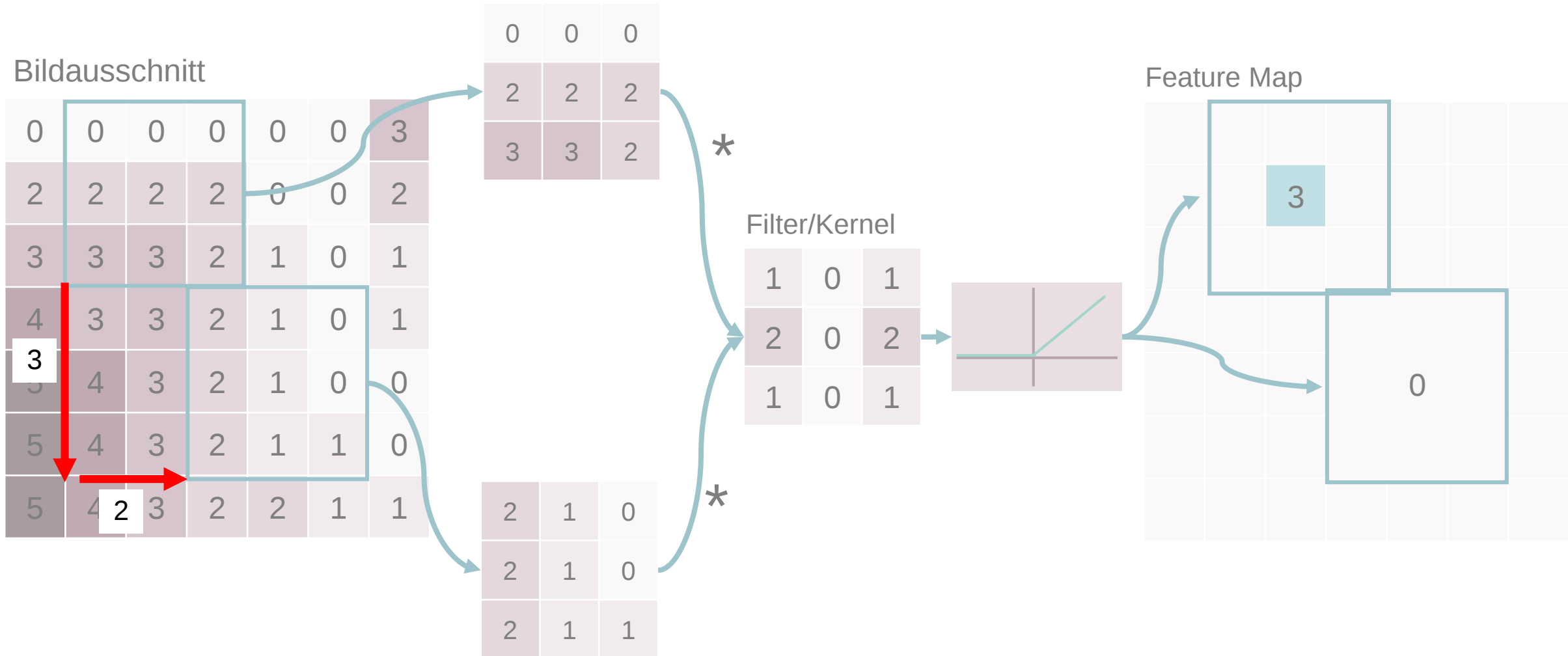
- Ziel: Quantitative Auswertung der Übereinstimmung
- Einstellung der gewünschten Übereinstimmung → Grenzwert
- Beurteilung der Überschreitung des Grenzwertes → Aktivierungsfunktion



# CNN: Räumliche Zuordnung II



# CNN: Abtastung des Bildes & Stride



# CNN: Abtastung des Bildes & Stride

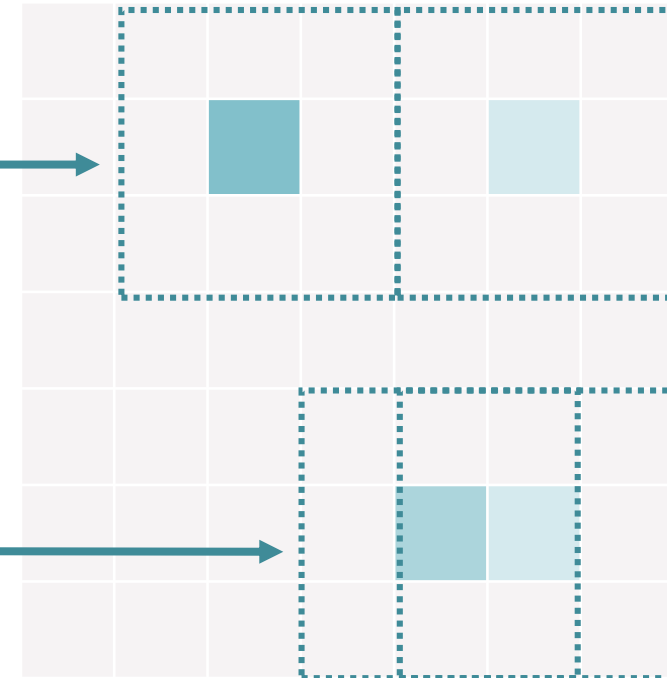
Bildausschnitt

|   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 3 | 0 | 0 | 3 |
| 2 | 2 | 2 | 2 | 0 | 0 | 0 | 2 |
| 3 | 3 | 3 | 2 | 1 | 0 | 1 | 1 |
| 4 | 3 | 3 | 2 | 1 | 0 | 1 | 1 |
| 5 | 4 | 3 | 2 | 1 | 0 | 0 | 0 |
| 5 | 4 | 3 | 2 | 1 | 1 | 1 | 0 |
| 5 | 4 | 3 | 2 | 2 | 1 | 1 | 1 |

Filter 3x3, Versatz (stride) 3  
(3x3 s3)

Filter 3x3, Versatz (stride) 1  
(3x3 s1)

Feature Map



Reduktion der  
räumlichen  
Auflösung

# CNN: Abtastung des Bildes & Stride

Bildausschnitt

|   |   |   |   |   |   |
|---|---|---|---|---|---|
| 0 | 0 | 2 | 2 | 0 | 3 |
| 2 | 2 | 2 | 0 | 0 | 2 |
| 3 | 3 | 2 | 1 | 0 | 1 |
| 3 | 3 | 2 | 1 | 0 | 1 |
| 4 | 3 | 2 | 1 | 0 | 0 |
| 4 | 3 | 2 | 1 | 1 | 0 |

Stride 3

(3x3 s3)

Feature Map



# CNN: Abtastung des Bildes & Stride

Bildausschnitt

|   |   |   |   |   |   |
|---|---|---|---|---|---|
| 0 | 0 | 2 | 2 | 0 | 3 |
| 2 | 2 | 2 | 0 | 0 | 2 |
| 3 | 3 | 2 | 1 | 0 | 1 |
| 3 | 3 | 2 | 1 | 0 | 1 |
| 4 | 3 | 2 | 1 | 0 | 0 |
| 4 | 3 | 2 | 1 | 1 | 0 |

Stride 3

Reduktion der  
räumlichen Auflösung

Feature Map

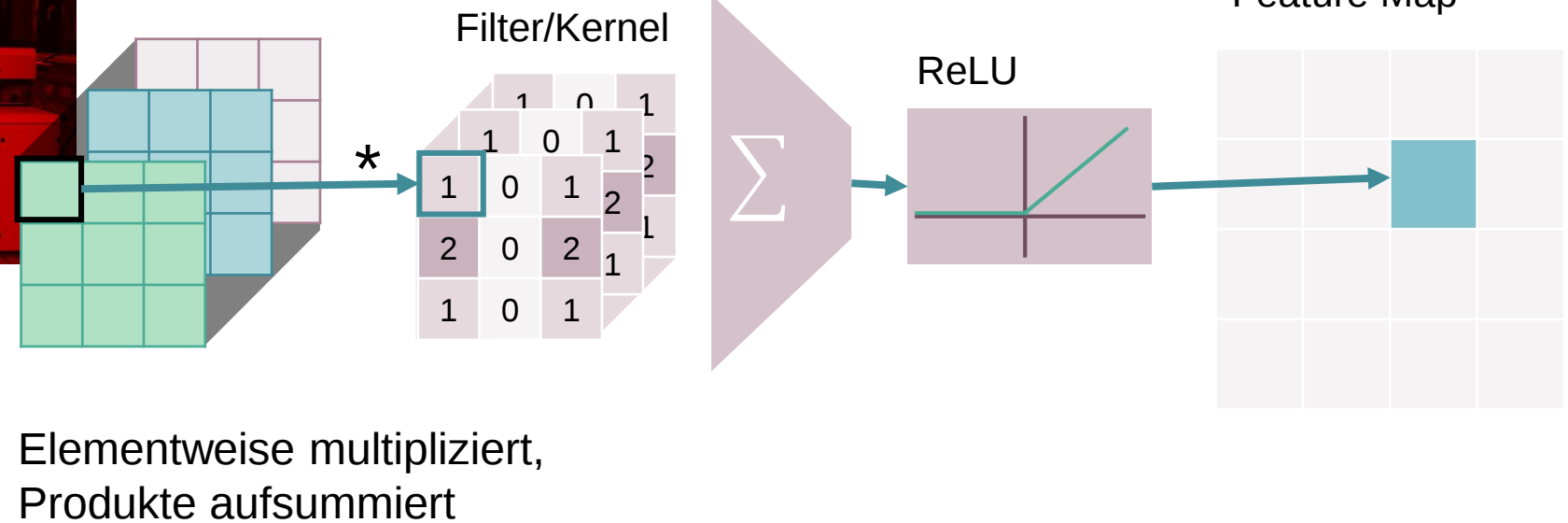
(3x3 s3)



Gewinn an  
semantischer  
Information

# CNN: Convolutions in Farbe

RGB Bild

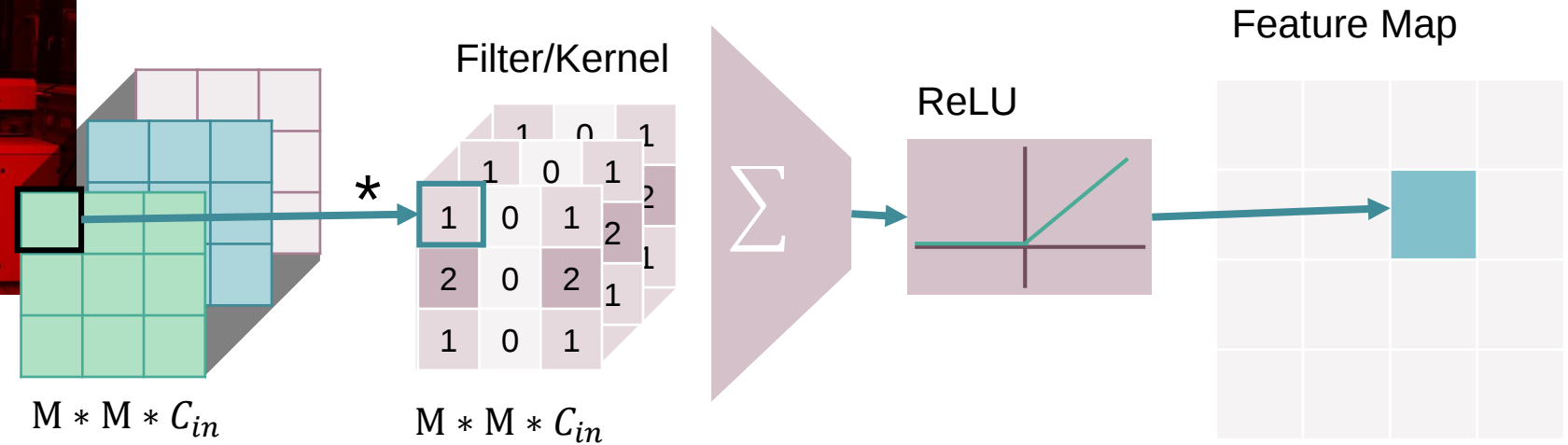


# CNN: Convolutions in Farbe

RGB Bild



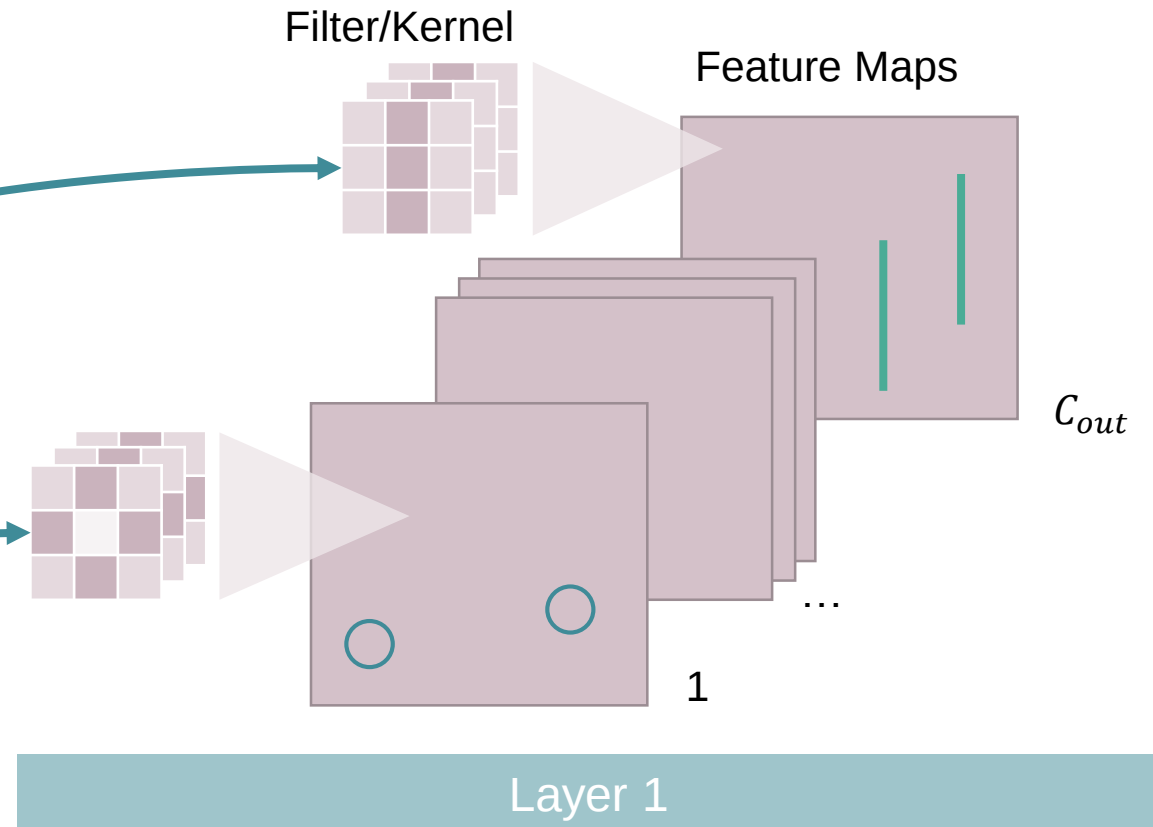
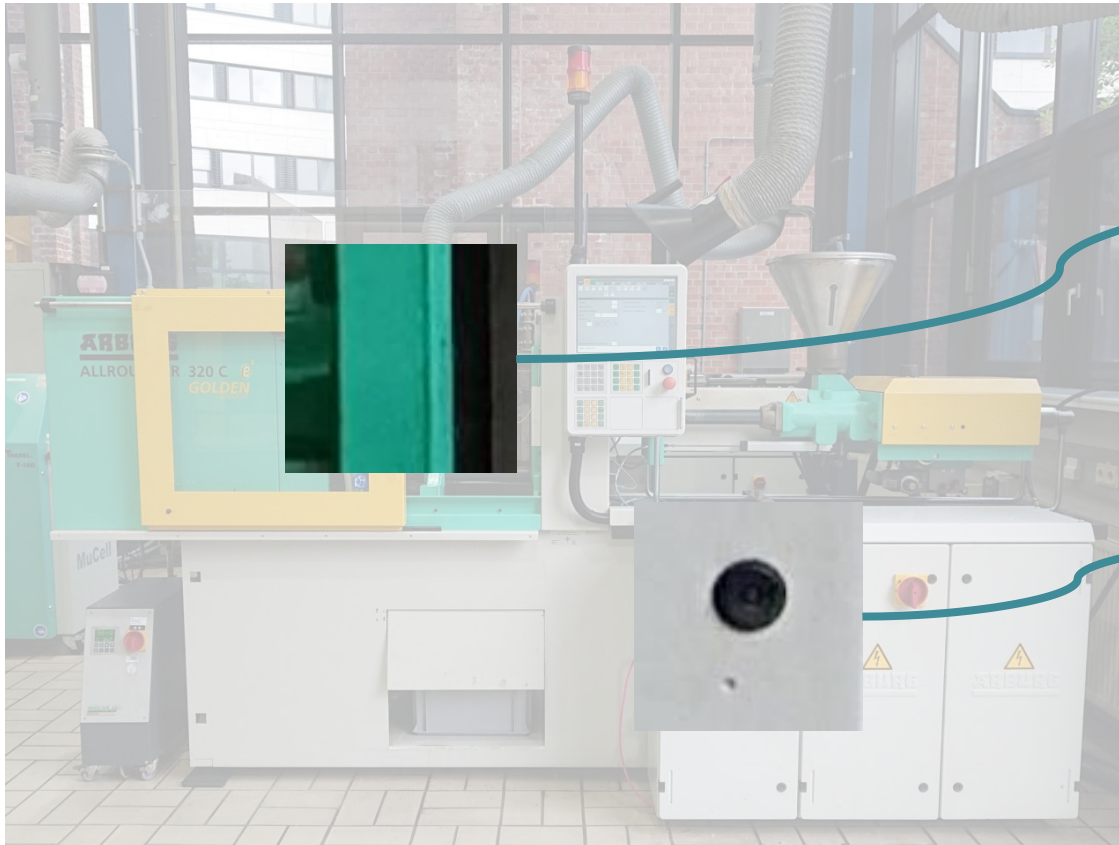
Höhe x Breite x 3



Parameter pro Feature Map:  $M^2 C_{in} + 1$

- Ein Grenzwert der Aktivierungsfunktion: +1
- Drei Farbkanäle RGB:  $C_{in} = 3$
- > Hier:  $3^2 3 + 1 = 28$

# CNN: Mehrere Features



Parameter pro Layer:  $(M^2 C_{in} + 1) * C_{out}$

# Welche Kernel/Filter gibt es?

Beispiele für Kernel

Identität

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Kantenerkennung

$$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix}$$

Nachschärfen

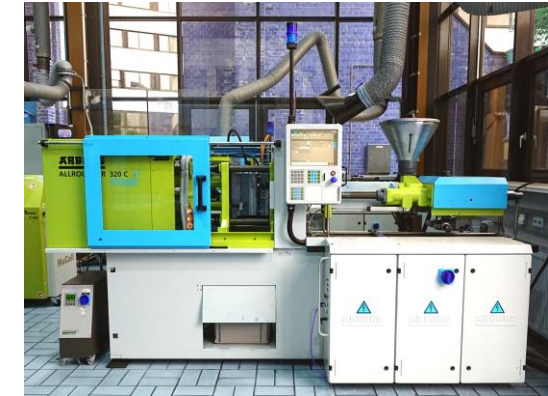
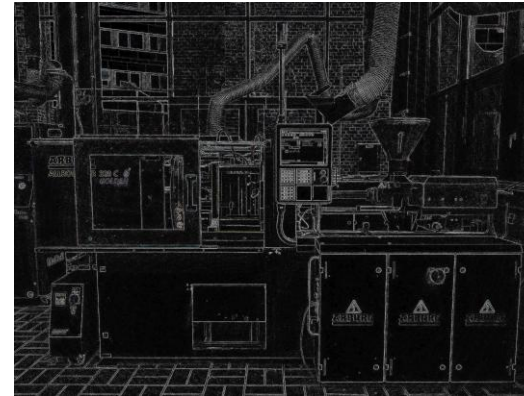
$$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$$

Weichzeichnen, Block

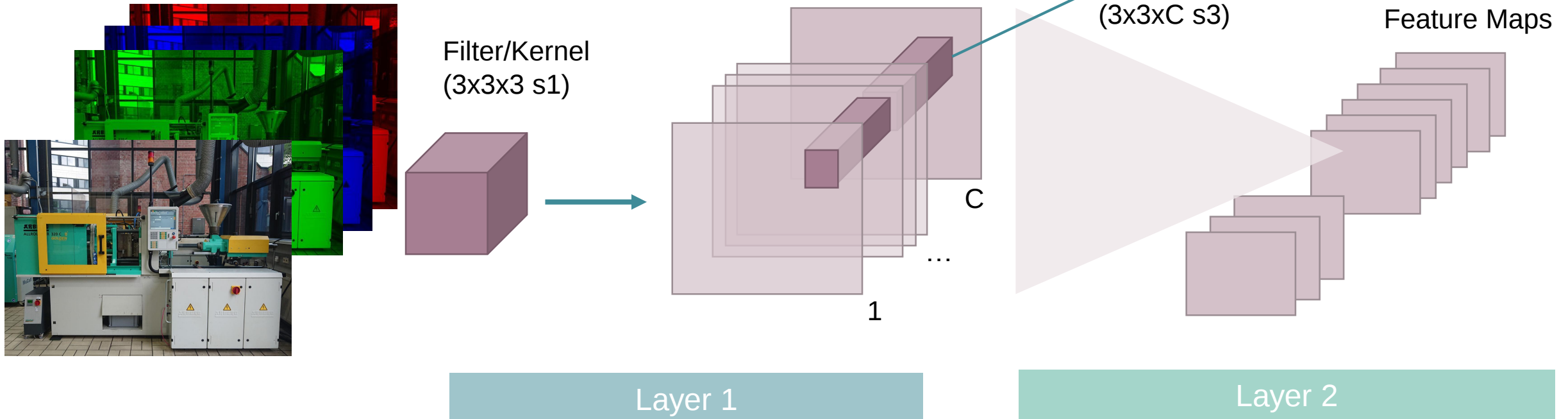
$$\frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Weichzeichnen, Gauß

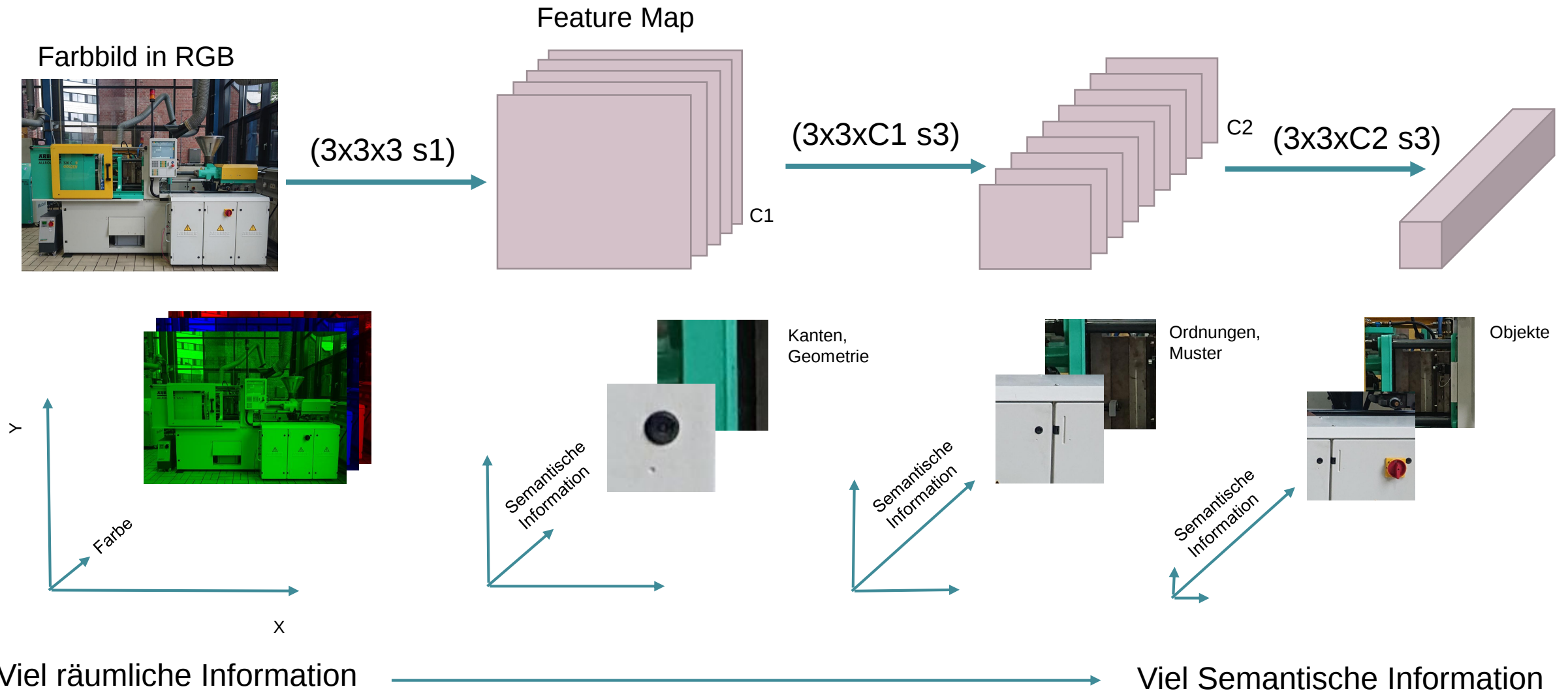
$$\frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$



# CNN: Hierarchische Muster



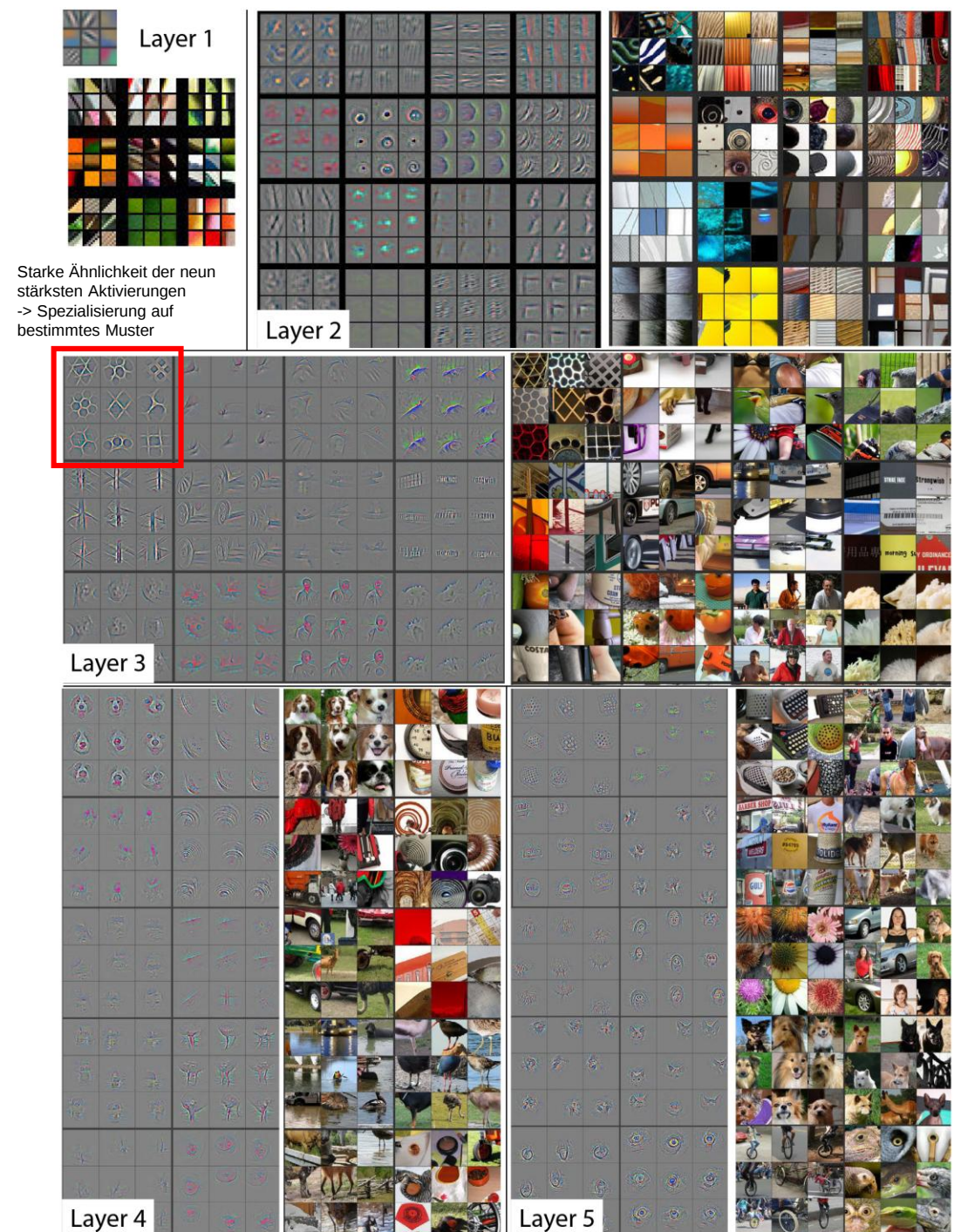
# CNN: Effekt mehrerer Schichten



# Zeiler und Fergus (2013)

## Visualizing and Understanding Convolutional Networks

„Visualization of features in a fully trained model. For layers 2-5 we show the top 9 activations in a random subset of feature maps across the validation data, projected down to pixel space using our deconvolutional network approach. Our reconstructions are not samples from the model: they are reconstructed patterns from the validation set that cause high activations in a given feature map. For each feature map we also show the corresponding image patches. Note: (i) **the the strong grouping within each feature map**, (ii) greater invariance at higher layers and (iii) exaggeration of discriminative parts of the image, e.g. eyes and noses of dogs.“ Zeiler und Fergus, 2013



# Zeiler und Fergus (2013)

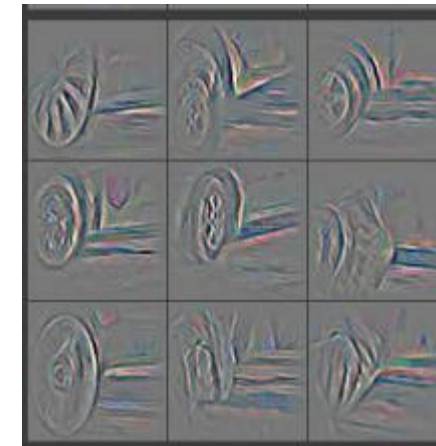
## Visualizing and Understanding Convolutional Networks

„Visualization of features in a fully trained model. For layers 2-5 we show the top 9 activations in a random subset of feature maps across the validation data, projected down to pixel space using our deconvolutional network approach. Our reconstructions are not samples from the model: they are reconstructed patterns from the validation set that cause high activations in a given feature map. For each feature map we also show the corresponding image patches. Note: (i) the strong grouping within each feature map, (ii) **greater invariance at higher layers and** (iii) **exaggeration of discriminative parts of the image**, e.g. eyes and noses of dogs.“ Zeiler und Fergus, 2013

Bsp.: Aktivierung der Feature Maps



Layer 1



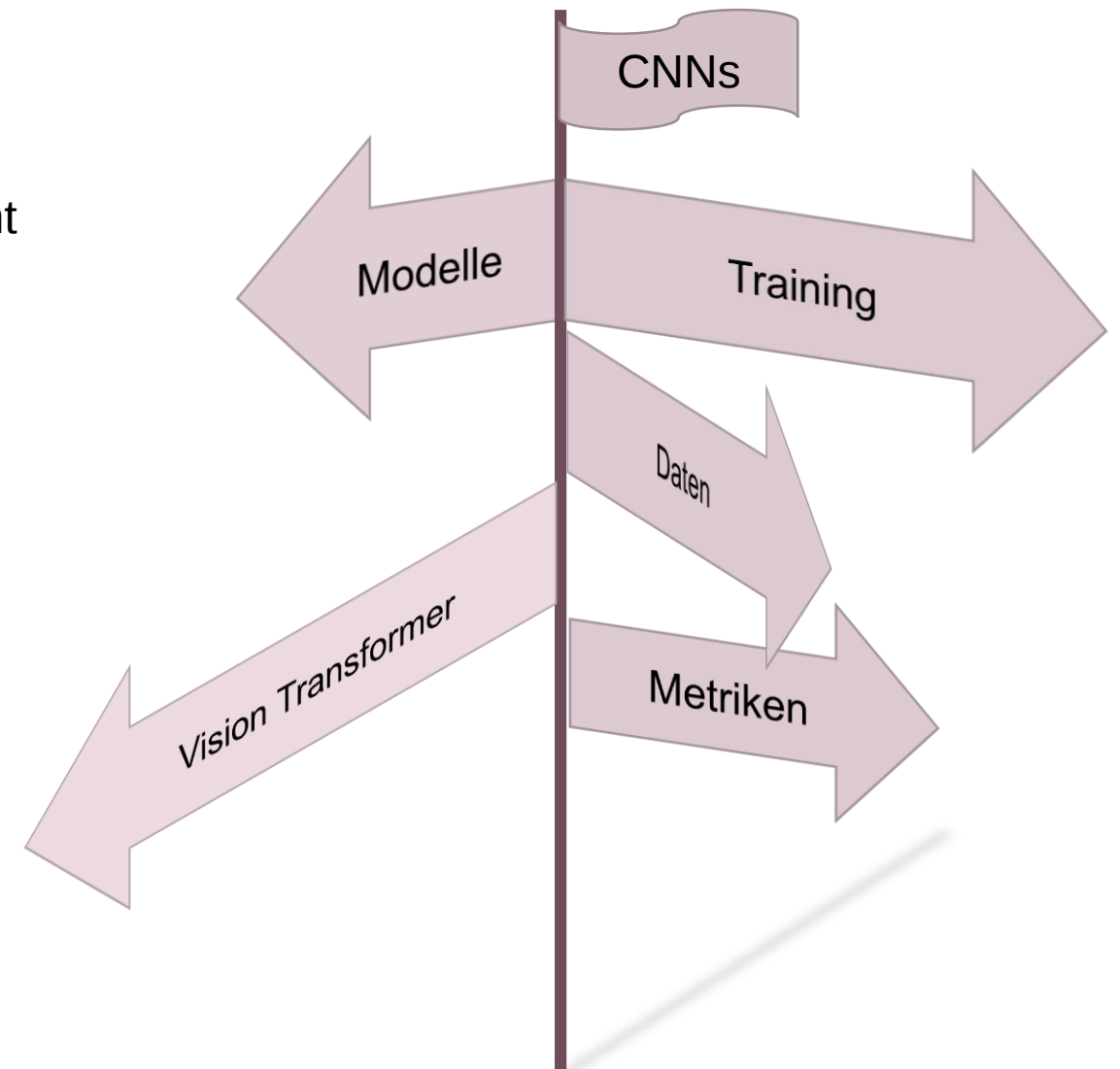
Layer 2



Layer 5

# Ausblick: Ein weites Feld

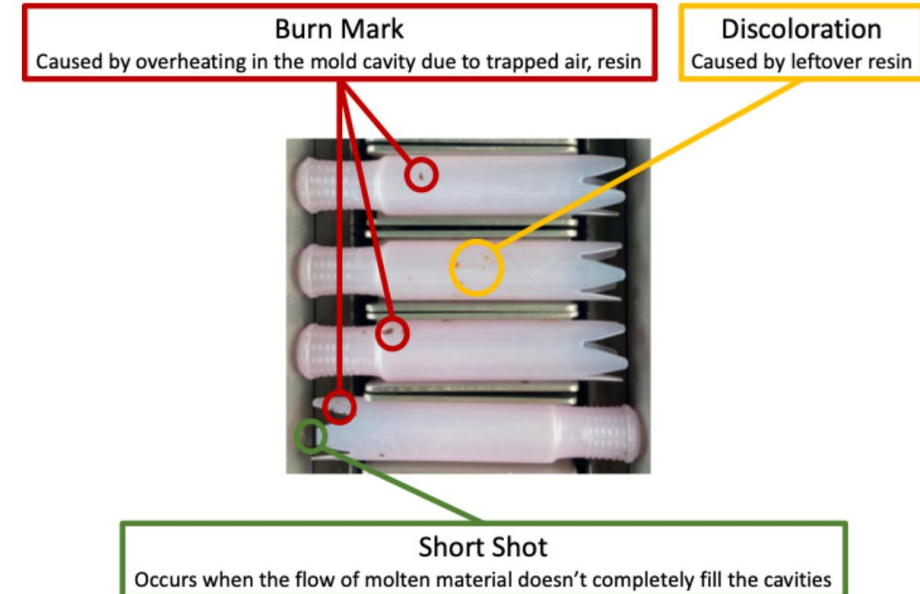
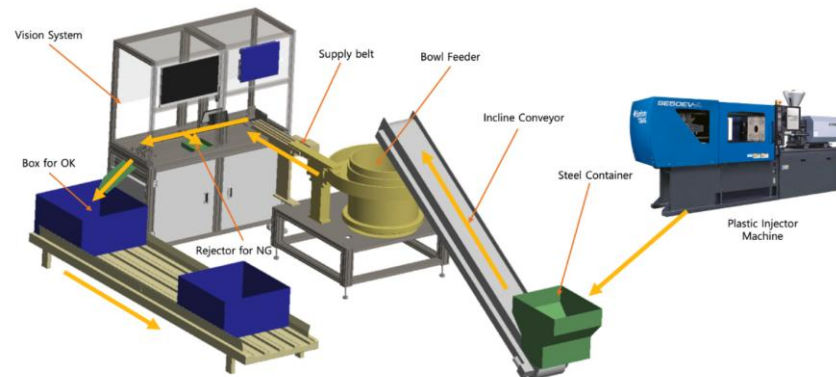
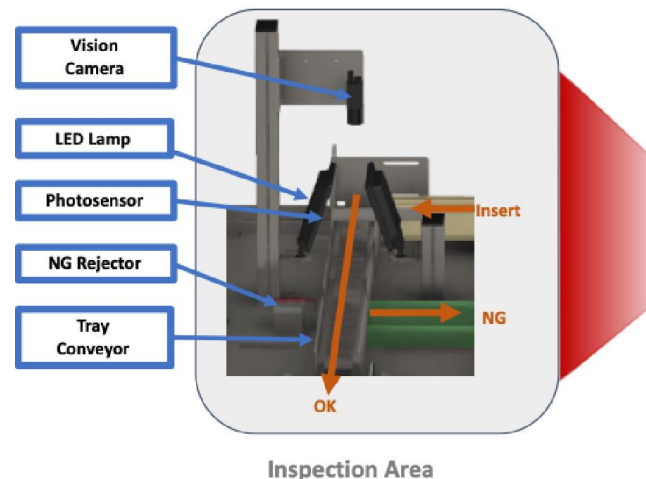
- Training  
Wie gewinnt man die optimalen Filter/Kernel?
- Daten  
Welche Rolle spielen die Bilddaten und woher bekommt man diese?
- Architektur/Modelle  
Welche Modelle gibt es in der Anwendung konkret und worin unterscheiden diese sich?
- Metriken  
Wie bewertet man die Qualität der Bilderkennung?
- Vision Transformer  
Die Zukunft der Bilderkennung?



# Aus der Praxis: Defekterkennung Spritzgießteilen

Im et al. (2021) A Data-Centric Approach to Design and Analysis of a Surface-Inspection System Based on Deep Learning in the Plastic Injection Molding Industry

- In-line Erkennung von Defekten an Produkten aus dem Hygienebereich
- Bilddatensatz von 150 Teile mit Defekten und unbekannte Zahl Gutteile aus der laufenden Produktionslinie
- Mehrfach fotografiert
- Annotation durch Expert:innen



Oben: Spritzgegossenes Hygieneprodukt mit betrachteten Defekt-Typen

Links: Aufbau zu in-line Datenerfassung und Bauteilsortierung

Abbildungen entnommen aus Im et al. (2021), Zusammenstellung angepasst

# Aus der Praxis: Defekterkennung Spritzgießteilen

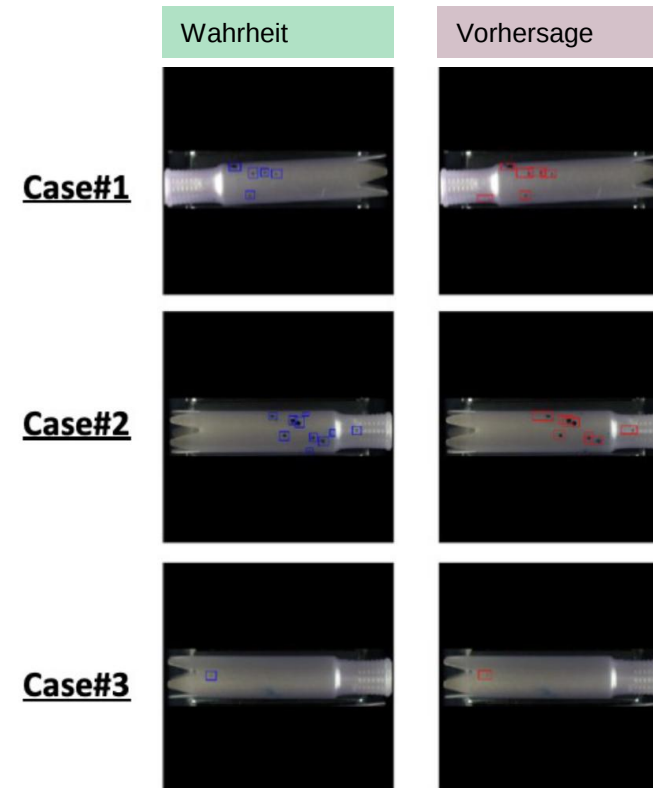
Im et al. (2021) A Data-Centric Approach to Design and Analysis of a Surface-Inspection System Based on Deep Learning in the Plastic Injection Molding Industry

- Objekterkennung mittels YOLOv4
- Überwacht trainierten
- *Recall* von 75 % (auf Bild-Level)
- Schluss: Annotation der Defekte Herausforderung

## Exkurs: Recall

Anteil der richtig erkannten, vorhandenen Defekten (TP) an allen, in Wirklichkeit vorhandenen Defekten (TP + FN)

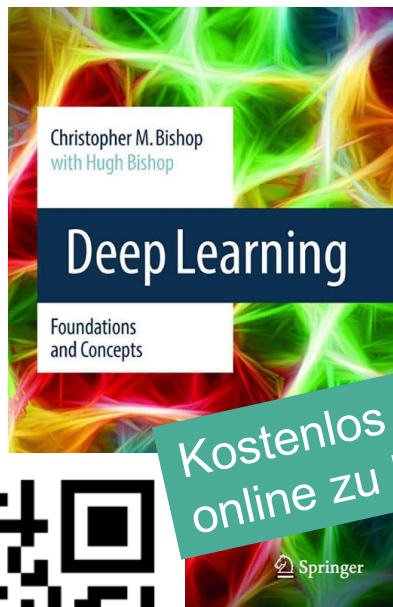
|          |                 | Vorhersage      |                                |
|----------|-----------------|-----------------|--------------------------------|
| Wahrheit | Richtig positiv | Falsch negativ  | $recall := \frac{TP}{TP + FN}$ |
|          | TP              | FN              |                                |
|          | Falsch positiv  | Richtig negativ |                                |
|          | FP              | TN              |                                |



Beispiele für fehlerhafte Erkennungen. Aus Im et al. (2021)

# Für interessierte Einsteiger und Fortgeschrittene

Deep Learning – Foundations and Concepts  
Bishop und Bishop (Springer, 2024)

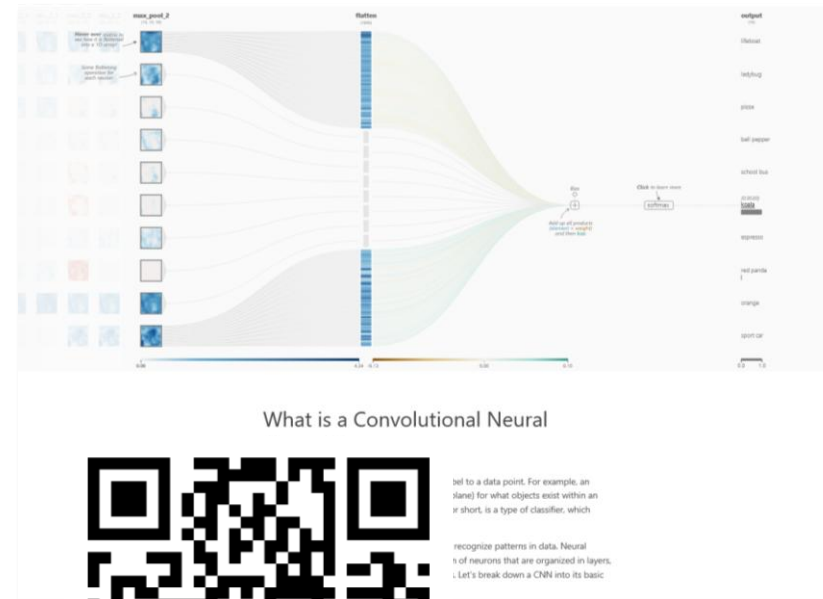


Kostenlos  
online zu lesen



<https://www.bishopbook.com/>

CNN Explainer: Learning Convolutional Neural Networks  
with Interactive Visualization, Wang et al. 2021



<https://poloclub.github.io/cnn-explainer/>



Digitale Lösungen zur Einführung von künstlicher  
Intelligenz in der Kunststoffverarbeitung

- Begegnet Herausforderung bei der digitalen Transformation in kleinen und mittelständischen Unternehmen (KMUs)
- Fokus: Relevante Prozessdaten für den Einsatz datengestützter Methoden erfassen, aufzubereiten und zielführend zu verwerten
- Praxisnahe Implementierung zweier Use-Cases:
  - Use-Case 1:  
Erkennung visueller Defekte im Spritzgießen
  - Use-Case 2:  
Prozessübergreifende Anomalien beim Rezyklateinsatz
- Umfangreiches und zielgruppenspezifisches Transferkonzept

## Weitere Informationen und Ansprechpartner

Oliver Kempf M. Sc.



+49 561 804-3599



kempf@uni-kassel.de



[www.uni-kassel.de/go/DUETT](http://www.uni-kassel.de/go/DUETT)



Kofinanziert von der  
Europäischen Union



Institut für Werkstofftechnik  
Kunststofftechnik  
Prof. Dr.-Ing. Hans-Peter Heim