

LRP für CNNs

- Basisformel

$$R_i^{(l)} = \sum_j \frac{z_{ij}}{\sum_{i'} z_{i'j}} R_j^{(l+1)} \quad \text{mit} \quad z_{ij} = x_i^{(1)} w_{ij}^{(l,l+1)}$$

- Problematisch bei komplexeren Layern (*Beispiel: ConvLayer*)
- Betrachte alternative Implementierung

LRP für CNNs

$$R_i^{(l)} = \sum_j \frac{z_{ij}}{\sum_{i'} z_{i'j}} R_j^{(l+1)} \quad \text{mit} \quad z_{ij} = x_i^{(1)} w_{ij}^{(l,l+1)}$$

- Algorithmus [QUELLE]

$$\begin{aligned} \forall_k : z_k &= \sum_{0,j} x_j \cdot \rho(w_{jk}) && \text{(Forward Pass)} \\ \forall_k : s_k &= R_k / z_k && \text{(Elementweise Division)} \\ \forall_j : c_j &= \sum_k \rho(w_{jk}) \cdot s_k && \text{(Backward Pass)} \\ \forall_j : R_j &= x_j c_j && \text{(Elementweises Produkt)} \end{aligned}$$

- Ergebnis identisch
- Was ist der Vorteil dieser neuen Reihenfolge?

LRP für CNNs

- Schritt 3 kann auch als Gradient ausgedrückt werden:

$$\forall_j : c_j = \sum_k \rho(w_{jk}) \cdot s_k \quad (\text{Backward Pass}),$$

- denn:

$$\begin{aligned} c_j &= \left[\nabla \left(\sum_k z_k(\mathbf{x}) \cdot s_k \right) \right]_j \\ &= \left[\nabla \left(\sum_k \left(\sum_{j'} (\mathbf{x}_{j'} \cdot w_{j'k}) \right) \cdot s_k \right) \right]_j \\ &= \left[\sum_k w_{jk} \cdot s_k \right]_j, \end{aligned}$$

wobei s_k als Konstante behandelt wird.

- Implementierung ist nun unabhängig der Art des Layers anwendbar.

Implementierte Varianten

■ LRP-0:

$$R_j = \sum_k \frac{a_j w_{jk}}{\sum_{0,j'} a_{j'} w_{j'k}} R_k$$

■ LRP- ϵ :

$$R_j = \sum_k \frac{a_j w_{jk}}{\epsilon + \sum_{0,j'} a_{j'} w_{j'k}} R_k$$

■ LRP- γ :

$$R_j = \sum_k \frac{a_j \cdot (w_{jk} + \gamma w_{jk}^+)}{\sum_{0,j'} a_{j'} \cdot (w_{j'k} + \gamma w_{j'k}^+)} R_k$$

Vergleichsbilder

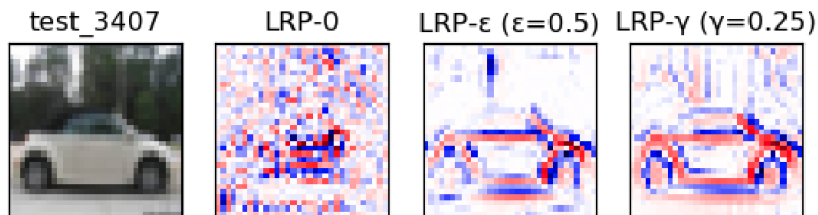


Abbildung: LRP-Varianten auf Cifar10-Datensatz

Vergleichsbilder

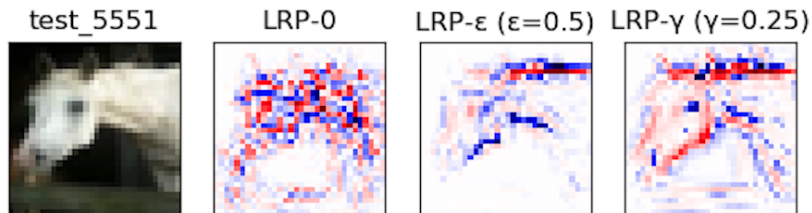


Abbildung: LRP-Varianten auf Cifar10-Datensatz

LRP-Composition

- **Idee:** Verwende unterschiedliche Varianten für verschiedene Layer des Netzes um Vorteile zu kombinieren
- **Beispiel:** Netz mit 10 Layern
 - LRP-0 auf Layern 10 und 9
 - LRP- ϵ auf Layern 8 bis 5
 - LRP- γ auf Layern 4 bis 1
- Auswahl der Parameter hängt stark von Netzarchitektur und der gewünschten Darstellung ab.

Beispiel LRP-Composition

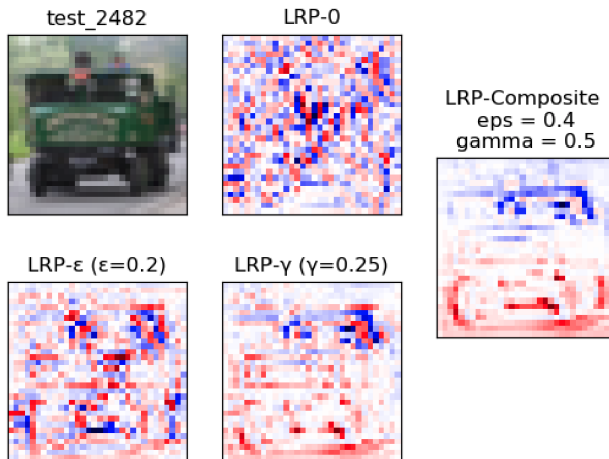


Abbildung: LRP-Composition auf Cifar10-Datensatz