Übungsblatt 7 : Fortgeschrittene Konvolutionsnetzwerke & Autoencoder
BA-INF 153: Einführung in Deep Learning für Visual Computing

Deadline:	**Theoretische Aufgaben**	26.06.2024 - 14:00 via eCampus
Tutoren:	Alina Pollehn	s6aapoll@uni-bonn.de
	Johannes van de Locht	s6jovand@uni-bonn.de
Übungsgruppenleitung:	Jan Müller	muellerj@cs.uni-bonn.de

Theoretische Übungsaufgaben (15 Punkte)

a) Back-propagation in Residual-Netzwerken *(4 Punkte)*

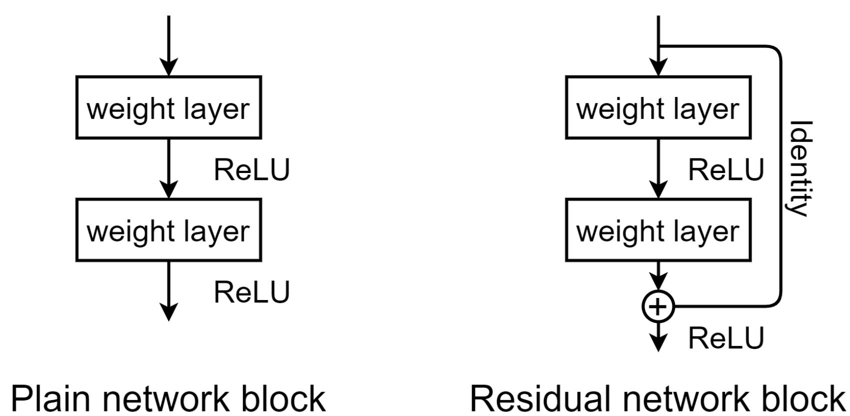

Figure 1: Eine Illustration von zwei Teilnetzblöcken. Links: Ein Block in einem einfachen Netz. Rechts: Ein Restblock mit zwei Schichten

Eine einfache Feedforward-Architektur, die wir in den vorangegangenen Übungen verwendet haben, modelliert die Funktion F, indem sie verschiedene Schichten nacheinander anwendet. Z.B.:

$$F(x) = h_n \text{ und } h_i = a_i(W_i \cdot h_{i-1}) \text{ und } h_0 = x$$

wobei a_i die Aktivierungsfunktion ist. Ein Residualnetz geht davon aus, dass F in einfache Funktionen F_i zerlegt werden kann und dass diese Funktionen als $F_i(x) = x + H_i(x)$ modelliert werden können, wobei $H_i(x)$ die Residualfunktion darstellt.

Ein einfaches Beispiel für ein Residualnetz ist

$$F(x) = h'_n \text{ und } h'_i = h'_{i-1} + a_i(W_i \cdot h'_{i-1}) \text{ und } h'_0 = x$$

wobei a_i auch hier die Aktivierungsfunktion ist. Abbildung 1 zeigt einen zweischichtigen Block eines einfachen Feed-forward Netzwerks und einen zweischichtigen Block eines Residualnetzwerks. **Sie können in dieser Aufgabe annehmen, dass a_i die Identitätsfunktion ist!**

- *(1 Punkt)* Berechnen Sie die Ableitung der Ausgabe h_n des einfachen Netzes nach einer vorherigen Schicht h_l, $n > l$.
- *(1 Punkt)* Berechne Sie die Ableitung der Ausgabe h'_n des Residualnetz nach einer vorhergehenden Schicht h'_l, $n > l$.
- *(2 Punkte)* Vergleichen Sie die beiden Ableitungen. Erläutern Sie anhand Ihres Vergleichs, warum tiefe Residualnetze erfolgreich trainiert werden können, während die Erhöhung der Tiefe eines einfachen Netzes sogar zu schlechteren Ergebnissen führen kann.

b) Autoencoder und Generative Modelle (*2 Punkte*)

Bisher haben wir uns ausschließlich mit diskriminativen Modellen befasst, da sie im Bereich der visuellen Erkennung die naheliegende Wahl sind. In der letzten Vorlesung wurden Autoencoder vorgestellt, die einen ersten Blick über diskriminative Modelle hinaus werfen. Um die grundlegenden Konzepte aus dieser Vorlesung zu festigen, sollten die folgenden Aspekte vertieft werden:

- *(2 Punkte)* Erläutern Sie die allgemeine Struktur eines Autoencoders. Erklären Sie, warum der Name "Autoencoder" für diese Art von Modell gewählt wurde. Erörtern Sie auch, ob ein Autoencoder ein generatives Modell ist, insbesondere in Bezug auf die Rolle des Encoders.

c) Unterschied zwischen zweier Verteilung messen *(4 Punkte)*

Die Kullback-Leibler-Divergenz (KLD) ist ein Maß für die Unterschiedlichkeit zweier Wahrscheinlichkeitsverteilung P und Q. Bei diskreten Verteilungen wird die KLD als Summe über die Ergebnismenge Ω berechnet. Sind P, Q reellwertige Verteilungen, wird die KLD als uneigentliches Integral über ihre Dichtefunktionen p und q berechnet:

KLD - Diskrete

$$D_{KL}(P \| Q) := \sum_{x \in \Omega} P(\{x\}) \log \left(\frac{P(\{x\})}{Q(\{x\})} \right)$$

KLD - Kontinuierlich

$$D_{KL}(P \| Q) := \int_{-\infty}^{\infty} p(x) \log \left(\frac{p(x)}{q(x)} \right) dx$$

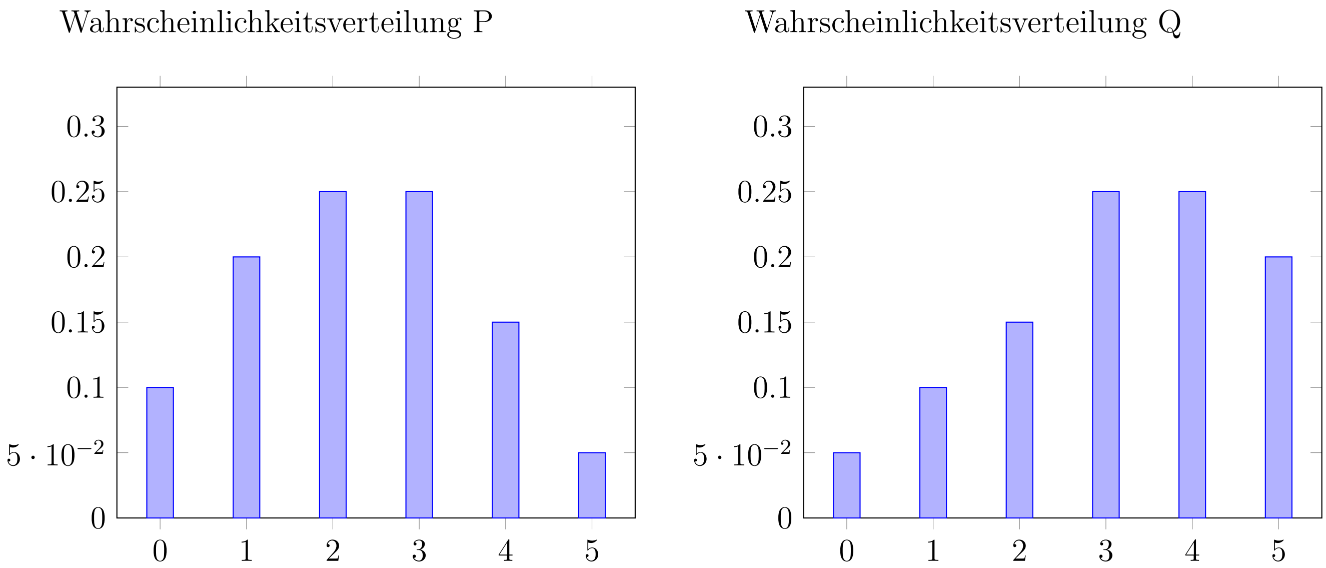

Figure 2: Histogramm der Wahrscheinlichkeitsverteilungen P and Q

Die KLD spielt eine große Rolle in der Zielfunktion eines Variational Autoencoder oder Diffusions Modelles. Wir wollen uns deshalb die KLD in den folgenden Aufgaben etwas genauer anschauen:

- *(2 Punkte)* Berechnen Sie die KLD zwischen zwei diskreten Wahrscheinlichkeitsverteilungen P und Q, deren Verteilung durch die Histogramme in Figur 2 gegeben ist.

- *(2 Punkte)* Eine Metrik ist die Verallgemeinerung unseres Verständnis für eine Distanz zwischen zwei Punkten und erfüllt die Eigenschaften: positiven Definitheit, Symmetrie und die Dreiecksungleichung. Zeige Sie, dass die Kullback-Leibler-Divergenz **keine** Metrik ist.

d) Geschlossene Form der KL-Divergenz zwischen Normalverteilung ***(5 Punkte)***

Die Kullback-Leibler-Divergenz zwischen zwei Normalverteilungen $\mathcal{N}(\mu_1, \Sigma_1)$ mit $\mathcal{N}(\mu_2, \Sigma_2)$ ist definiert als

$$D_{KL}(\mathcal{N}(\mu_1, \Sigma_1) \parallel \mathcal{N}(\mu_2, \Sigma_2)) = \int_{\mathbf{h}} f(\mathbf{h} \mid \mu_1, \Sigma_1) \log\left(\frac{f(\mathbf{h} \mid \mu_1, \Sigma_1)}{f(\mathbf{h} \mid \mu_2, \Sigma_2)}\right) \mathrm{d}\mathbf{h}.$$

wobei f die Dichtefunktion der Normalverteilung mit Mittelwert $\mu \in \mathbb{R}^n$ und Kovarianzmatrix $\Sigma \in \mathbb{R}^{n \times n}$ ist

$$f(\mathbf{h} \mid \mu, \Sigma) = \frac{1}{(2\pi)^{n/2} \det(\Sigma)^{1/2}} e^{-\frac{1}{2}(\mathbf{h}-\mu)^T \Sigma^{-1} (\mathbf{h}-\mu)}.$$

Im Variational Autoencoder beschränken wir uns auf den Spezialfall in dem $\Sigma_1 = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^2)$ eine Diagonalmatrix ist und $N(\mu_2, \Sigma_2)$ die Standartnormalverteilung in n Dimensionen ist d.h. $\mu_2 = \mathbf{0}$ und $\Sigma_2 = \mathbf{I}$. Zeigen Sie, dass die KL-Divergenz in diesem Spezialfall durch die folgende geschlossene Form berechnet werden kann

$$D_{KL}\left(N(\mu_1, \mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^2)) \parallel N(\mathbf{0}, \mathbf{I})\right) = \frac{1}{2}\left[-\sum_i (\log(\sigma_i^2) + 1) + \sum_i \sigma_i^2 + \sum_i \mu_i^2\right].$$

Hinweis: Verwenden Sie ohne Beweis, dass $D_{KL}(P \parallel Q) = E_P[\log P - \log Q]$ und dass die Quadratik $x^T Ax$ bei einer positiv-definiten Matrix A den Erwartungswert $E_P[x^T Ax] = E_P[x]^T AE_P[x] + \mathrm{tr}(A\Sigma)$ besitzt.