

---

## 13 Style Transfer

Systemtechnik BSc  
FS 2025

---

### Aufgaben

Applied Neural Networks im Modul ANN | WUCH

---

### Lernziele

Nach dem Bearbeiten dieser Übungsserie ...

- verstehen Sie, dass ein CNN in der Lage ist, Inhalt (content) und Stil, Textur (style) eines Bildes zu trennen, und wie mittels Neural Style Transfer der Inhalt eines Bildes mit der Textur eines anderen Bildes (Style) gemischt werden kann.  
Aufgaben: 1
- verstehen Sie, dass das vortrainierte VGG19-Netz für die Technik des Style-Transfers nur dazu dient, die Trennung zwischen Style und Content vorzunehmen und die entsprechenden Loss-Funktionen zu berechnen, und dass das generierte Bild  $X$  ähnlich berechnet wird wie Inputs bei XAI (Gradientenabstieg nach  $X$  anstelle von  $w$ ).  
Aufgaben: 1
- können Sie Google Colab verwenden, um rechenaufwändige Deep-Learning Methoden zu testen.  
Aufgaben: 1

### Neural Style Transfer

Neural Style Transfer ist ein Verfahren, bei dem der Inhalt ( $C$ ) eines Bildes mit einem Stil eines weiteren Bildes ( $S$ ) kombiniert werden kann. Als Stil-Bilder eignen sich Kunstgemälde sehr gut, da diese deutlich erkennbare, globale Muster und Texturen enthalten. Das Verfahren benötigt ein künstliches neuronales Faltungsnetzwerk (CNN), hier ein

VGG19, welches die Bildanalyse vornimmt. Diese Bildanalyse ist dazu da, um getrennt Informationen über den *Inhalt* und den *Stil* der Bilder zu extrahieren, damit diese anschliessend miteinander verglichen werden können.

Convolutional Neural Networks (CNN) sind mehrschichtige künstliche neuronale Netze, welche für die Bildanalyse oder Kategorisierung von Bildern verwendet werden. Die CNNs bestehen mehrheitlich aus Convolutional Layern, welche anhand von Filtern bestimmte Muster, wie Linien oder Kreise, erkennen. Da die Dimensionen der Filterkerne (kernel) relativ klein sind, verfügen sie über ein beschränktes *Wahrnehmungsfeld* (field of reception) und decken so nur einen kleinen Bildbereich ab. Durch *Pooling* Layer kann die Bildgrösse verkleinert werden und die Objektinformationen rücken näher zusammen, wodurch Muster mit grösserer Längenskala erkannt werden. **In den vorderen Schichten werden kleinreichweitige Muster (Texturen, Stil) erkannt werden und in den hinteren Schichten grossreichweitige Muster (Struktur, Inhalt).**

Da für die Stil-Extraktion klein- und grossreichweitige Muster benötigt werden, werden Informationen von mehreren Schichten verwendet. Der Inhalt wird in hinteren Schichten ausgelesen, da dort das Gesamt-Bild erkennbar wird.

Diese Übungsserie zeigt, wie Sie mit PyTorch auf Google Colab ihre eigenen Bilder mit selbst gewählten Mustern stilisieren können.

### Content- und Style-Loss

Um den Style Transfer durchzuführen, werden drei Bilder ( $S, C, X$ ) benötigt. Das erste Bild  $S$  enthält den gewünschten Stil (Beispiel Sternennacht von Van Gogh), das zweite enthält den Inhalt (Content  $C$ , z.B. Bild der Neckarfront in Tübingen) und beim dritten handelt es sich um das generierte Bild  $X$  aus Stil-Bild und Inhalt-Bild. Das generierte Bild wird zu Beginn mit dem Inhalt-Bild initialisiert. Das Ziel besteht darin, dass das generierte Bild den Inhalt vom Inhalt-Bild behält, dies jedoch im Stil vom Stil-Bild angezeigt wird. Dies wird realisiert, indem anhand des VGG19 der Stil des Stil-Bildes und des generierten Bildes ausgelesen und miteinander verglichen wird. Aus der berechneten Differenz ergibt sich der Fehler des Stils (Style-Loss  $\mathcal{L}_{\text{style}}$ ). Dasselbe wird beim Inhalt gemacht und ebenfalls miteinander verrechnet (Content-Loss  $\mathcal{L}_{\text{content}}$ ).

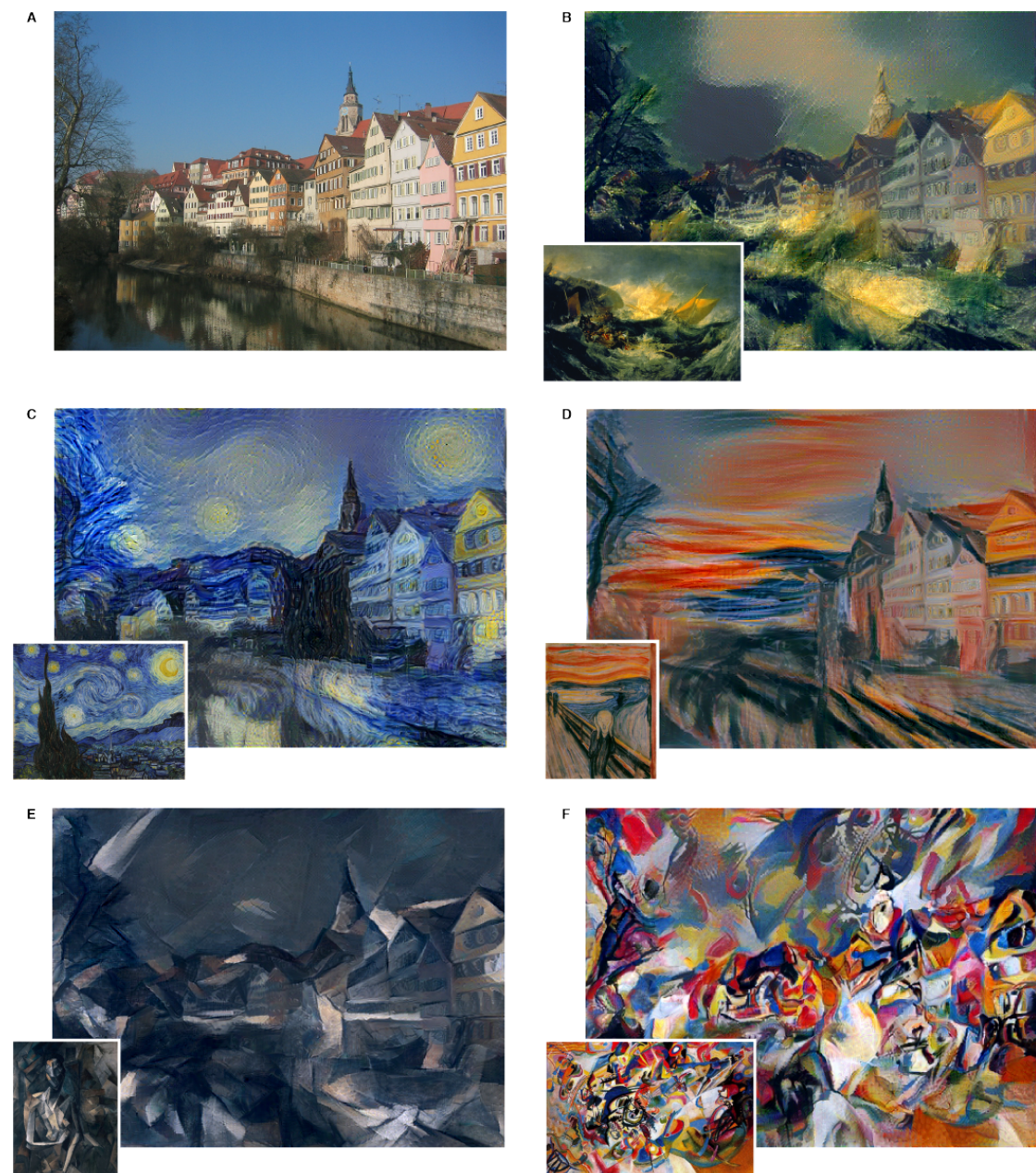


Abbildung 1: Bilder, die den Inhalt einer Fotografie mit dem Stil mehrerer bekannter Kunstwerke kombinieren. **A** *Neckarfront in Tübingen, Germany* Foto von Andreas Praefcke **B** *The Shipwreck of the Minotaur* von J.M.W. Turner, 1805. **C** *The Starry Night* von Vincent van Gogh, 1889. **D** *Der Schrei* von Edvard Munch, 1893. **E** *Femme nue assise* vom Pablo Picasso, 1910. **F** *Composition VII* von Wassily Kandinsky, 1913.

$$\mathcal{L}_{\text{total}}(\vec{p}, \vec{a}, \vec{x}) = \alpha \mathcal{L}_{\text{content}}(\vec{p}, \vec{x}) + \beta \mathcal{L}_{\text{style}}(\vec{a}, \vec{x}) \quad (1)$$

Mittels Gradientenverfahren wird der Gesamtfehler  $\mathcal{L}_{\text{total}}$  aus von Inhalt und Stil-Loss minimiert.

### Aufgabe 1. Neural Style Transfer

- (a) Öffnen Sie das Notebook ANN13\_neural\_style\_transfer\_TEMPLATE\_pl.ipynb. Sie können es z. B. in Google Colab ausführen, indem Sie auf den entsprechenden Link oben im Notebook klicken.

**Setup-Funktion:** Untersuchen Sie die Funktion `setup_environment()`. Welche Dateien werden heruntergeladen und in welchen Ordnern gespeichert? Wofür werden die VGG-Gewichte und die Bilder später verwendet?

- (b) **Hyperparameter:** In Zelle (b) werden wichtige Parameter definiert. Welche Parameter steuern die Gewichtung von Content, Style und Bildglättung? Welche Auswirkungen könnte eine Veränderung dieser Parameter haben?
- (c) **Bildverarbeitung:** Erläutern Sie die Funktionen `load_image`, `show`, `saveimg` und `transfer_color`. Welche Aufgabe erfüllt die Farbübertragung (`PRESERVE_COLOR`)?
- (d) **Tensor-Konvertierung:** Untersuchen Sie die Funktionen `itot` und `ttot`. Warum werden Bilder für VGG19 normalisiert? Welche Werte werden verwendet?
- (e) **Feature Extraction:** Wie funktioniert die Funktion `get_features(model, tensor)`? Welche Layer werden für den Content-Loss bzw. Style-Loss verwendet? Welche Rolle spielt die Gram-Matrix?
- (f) **Loss-Funktionen:** Vergleichen Sie die Definitionen von `content_loss`, `style_loss` und `tv_loss` mit der Darstellung im Paper von Leon A. Gatys et al. (2015). Wofür steht der TV-Loss und warum wird er häufig eingesetzt?
- (g) **Initialisierung:** Die Funktion `initial()` erzeugt das zu optimierende Bild. Was ist der Unterschied zwischen "random" und "content"? Wie wirkt sich die Wahl auf das Ergebnis aus?

- (h) **Optimierung:** Die Funktion `stylize()` führt den eigentlichen Style Transfer durch. Welche Verluste werden kombiniert? In welchen Intervallen wird das Bild angezeigt und gespeichert? Welche Rolle spielt die Closure-Funktion bei der Optimierung?
- (i) **Visualisierung:** Verwenden Sie die Funktion `plot_losses()` und interpretieren Sie den Verlauf der einzelnen Verlustkomponenten (Content, Style, TV, Total). Welche Komponenten dominieren den Gesamtverlust?
- (j) **Eigene Bilder verwenden:** Ersetzen Sie die beiden Bilder in `data/images/` durch eigene Content- und Style-Bilder Ihrer Wahl (achten Sie auf sinnvolle Auflösung und Format). Führen Sie den Style Transfer erneut aus. Beurteilen Sie die Ergebnisse. Variieren Sie ggf. Hyperparameter wie `STYLE_WEIGHT` und `CONTENT_WEIGHT`.
- (k) **Zusatz: Vergleich mit Fast Style Transfer** (optional): Lesen Sie im Notebook den Abschnitt zur alternativen Methode des „Fast Style Transfer“. Erläutern Sie den Unterschied zur hier verwendeten Optimierung des Bildes selbst. Welche Vor- und Nachteile bietet ein trainiertes Stilnetz?





Abbildung 2:

## Lösungen

### Lösung 1.

Sie finden die Lösung auf Moodle unter `ANN13_neural_style_transfer_SOLUTION_p1.ipynb`.