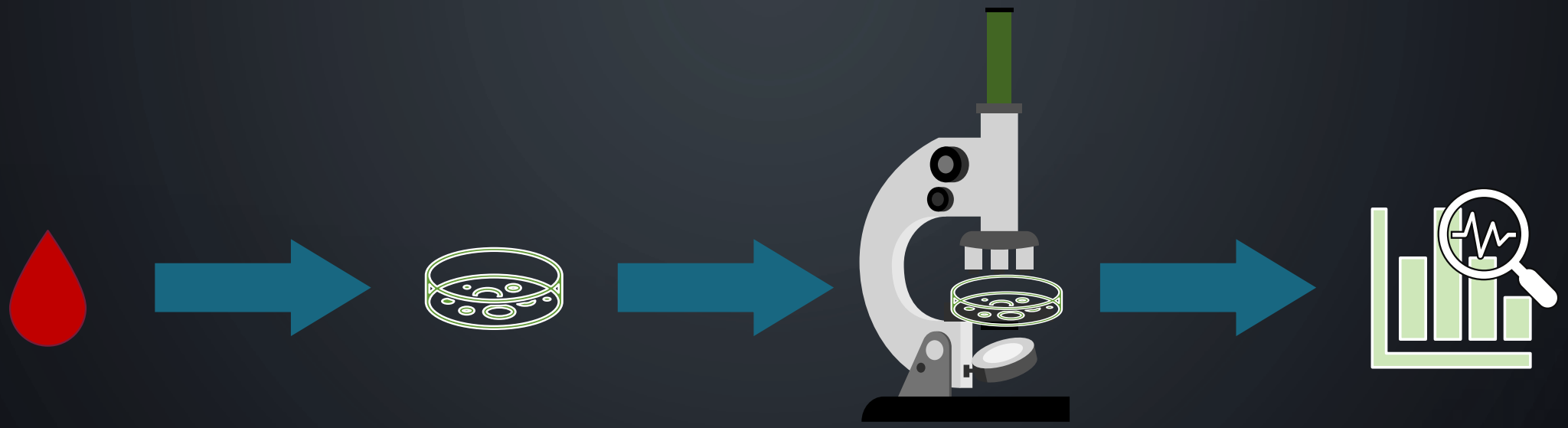
“Explainable Artificial Intelligence“ für Stammzellmodelle des Leigh-Syndroms

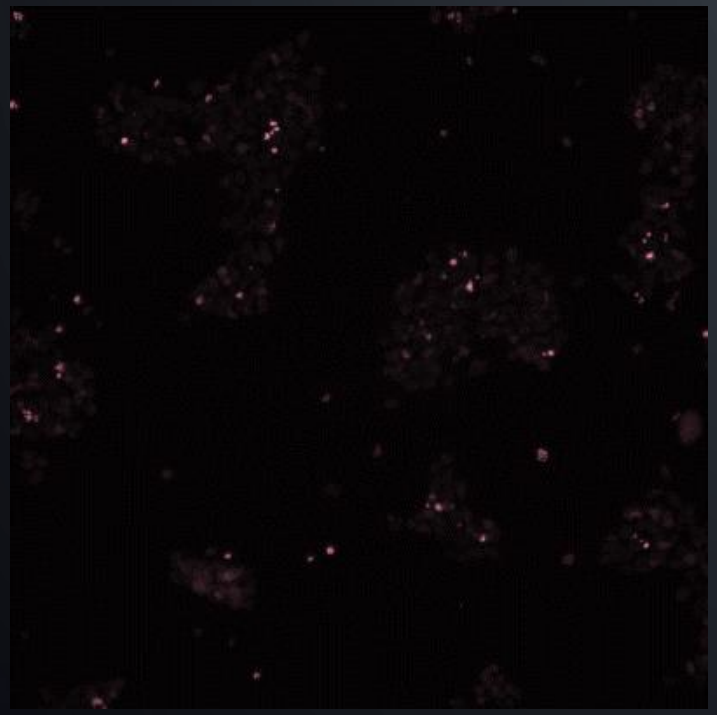
Maximilian Otto

Erstprüfer: Prof. Dr. Landgraf (FU Berlin)

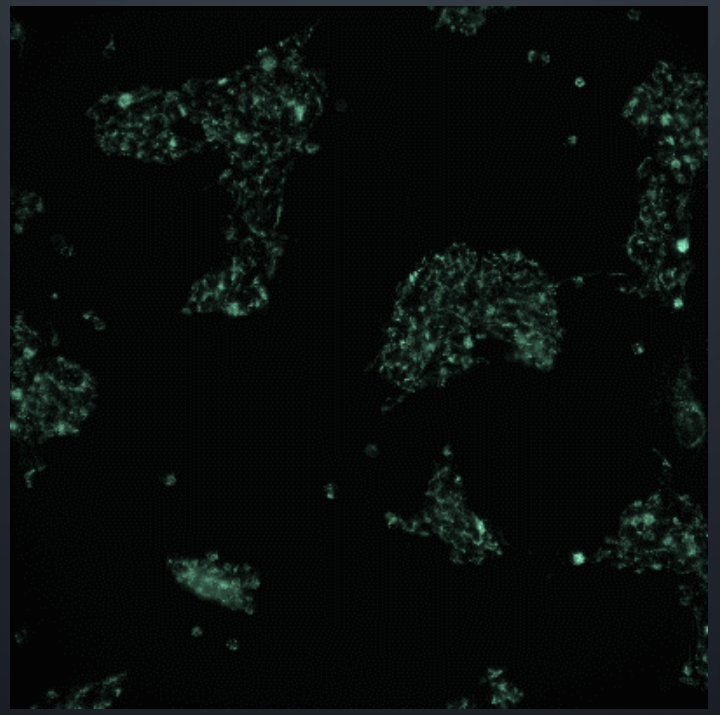
Betreuer: Dr. Metzger (MDC)



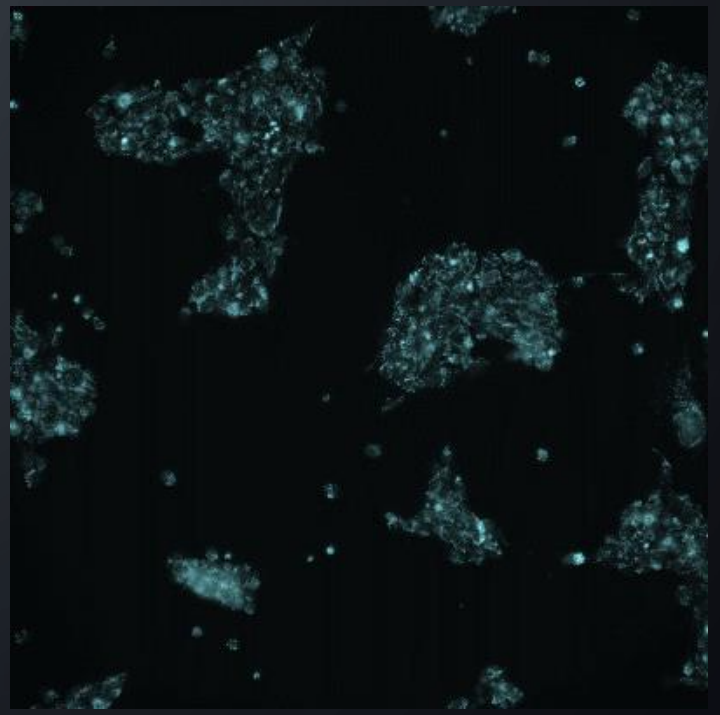
ch1



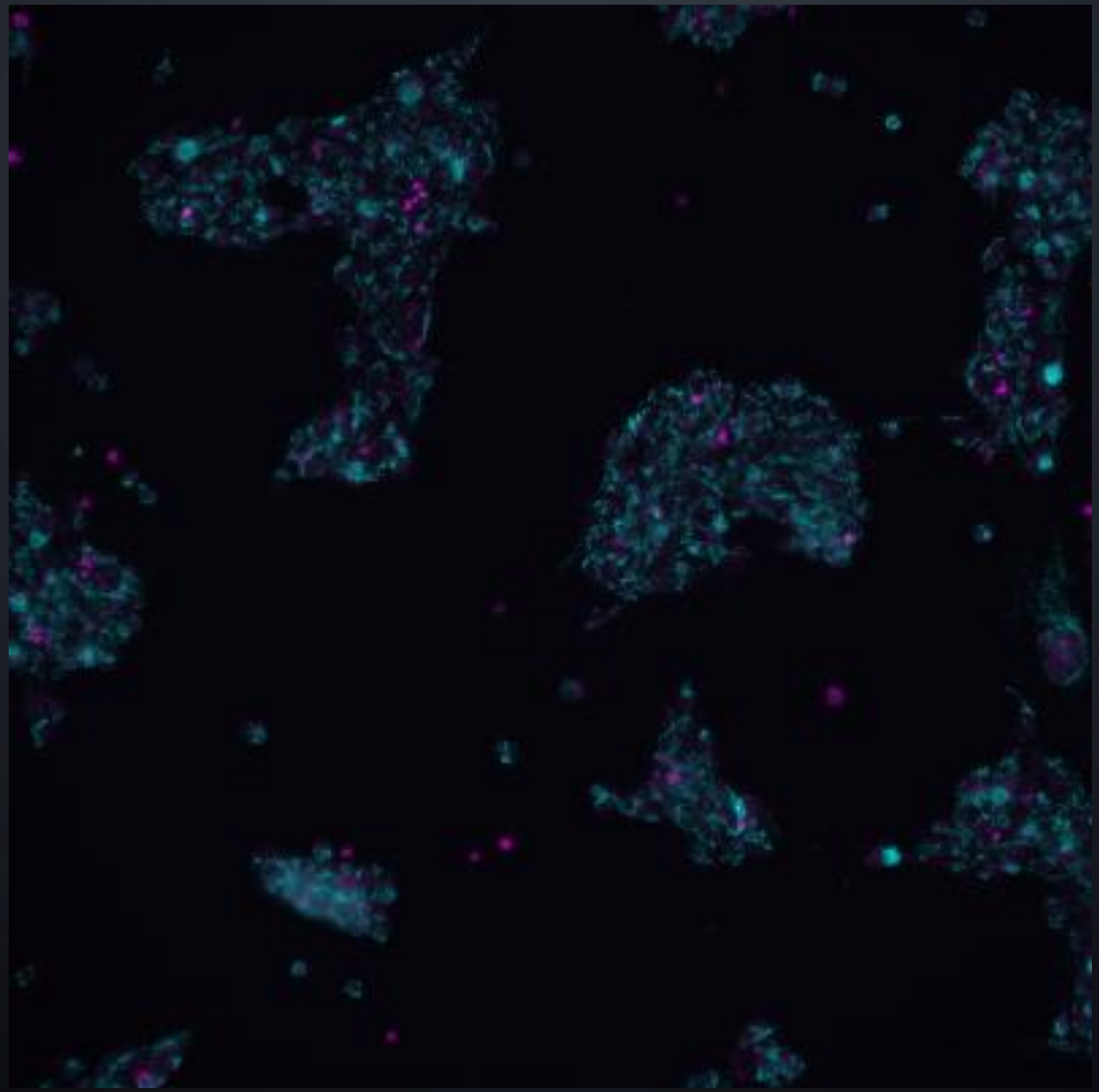
ch2



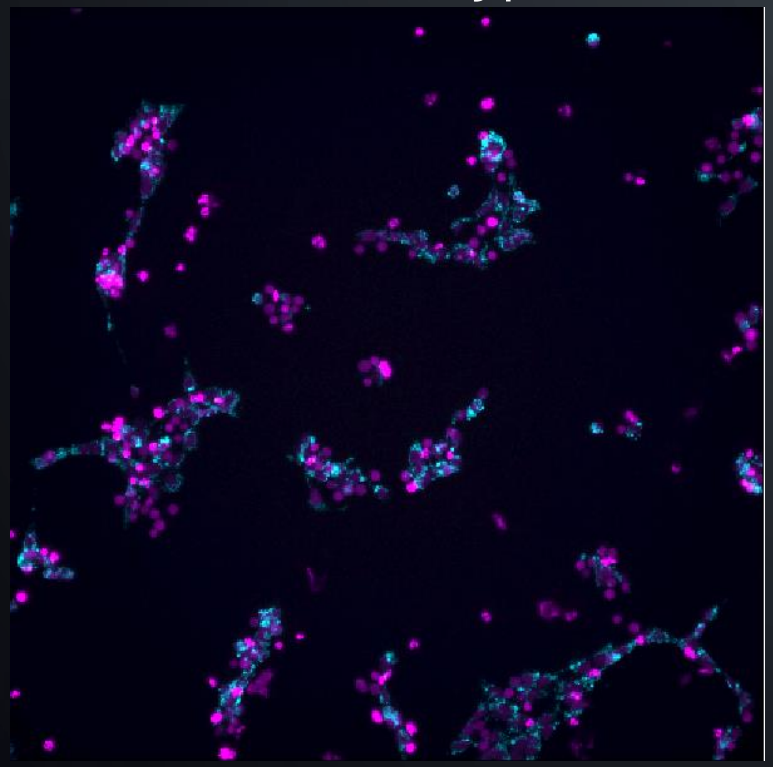
ch3 (ch2 + ch1)



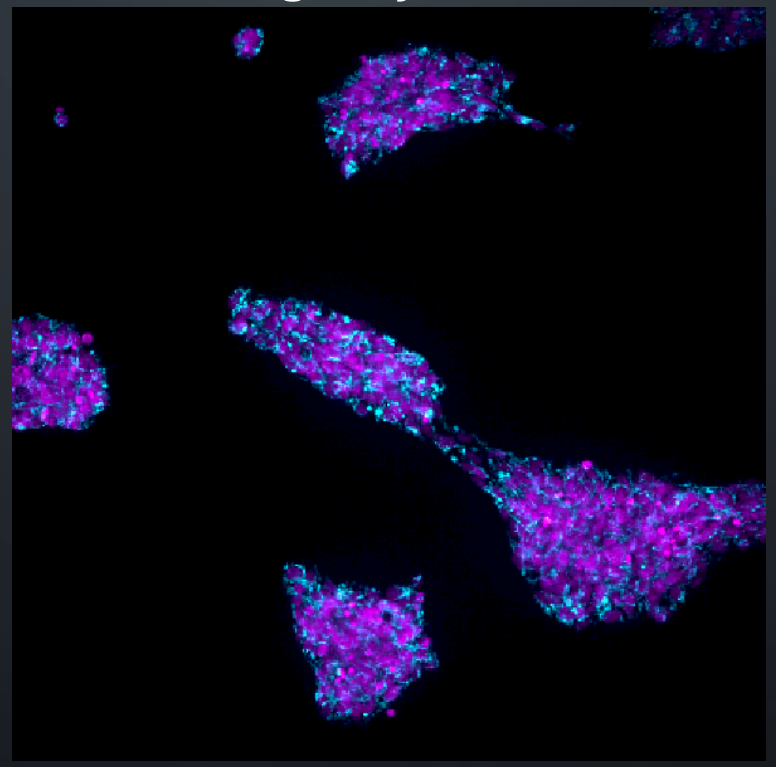
Kombiniert:



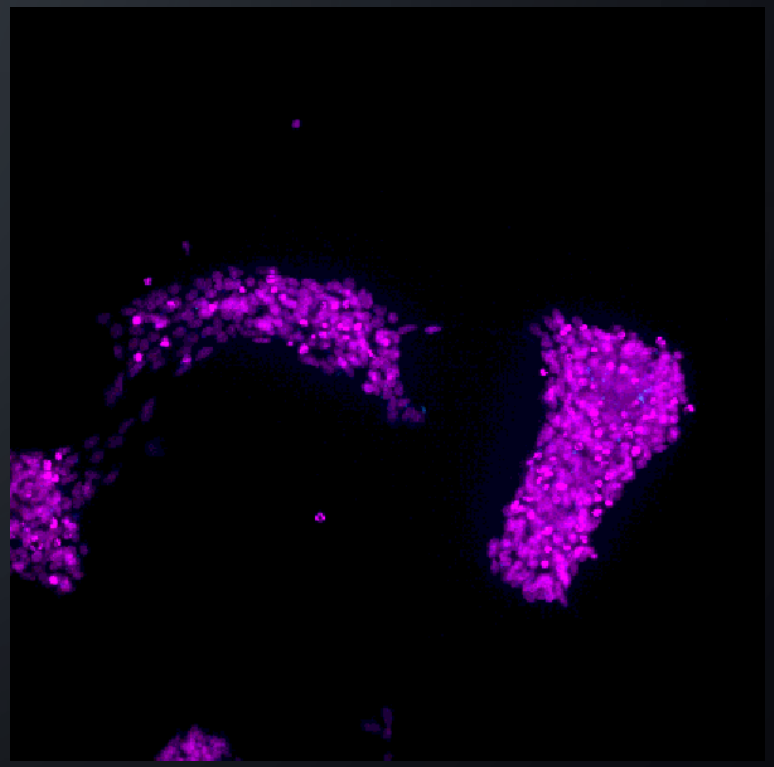
Wildtyp



Leigh-Syndrom



Behandelt

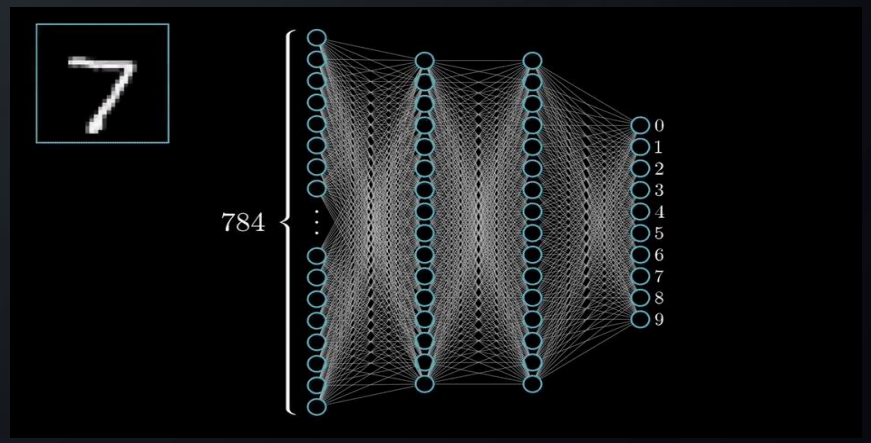


KLASSIFIKATION MIT RESNEXT-50 (DS VS. TT)

540 von 540 korrekt klassifiziert

	precision	recall	f1-score	support
DS	1.00	1.00		264
TT	1.00	1.00		276
accuracy			1.00	540

Z'-Factor: 0.924



3Blue1Brown, „Chapter 1 Deep Learning“, 05.10.2017

WAS KÖNNEN WIR DAMIT MACHEN?

> Früherkennung von Krankheiten unter kontrollierten Umständen

> Wirkstoffe für eine mögliche Behandlung evaluieren

-> Welche Kulturen lassen sich nach Behandlung wie gut von normalen oder erkrankten unterscheiden?

> Versuchen, die Unterschiede zwischen den Klassen zu visualisieren

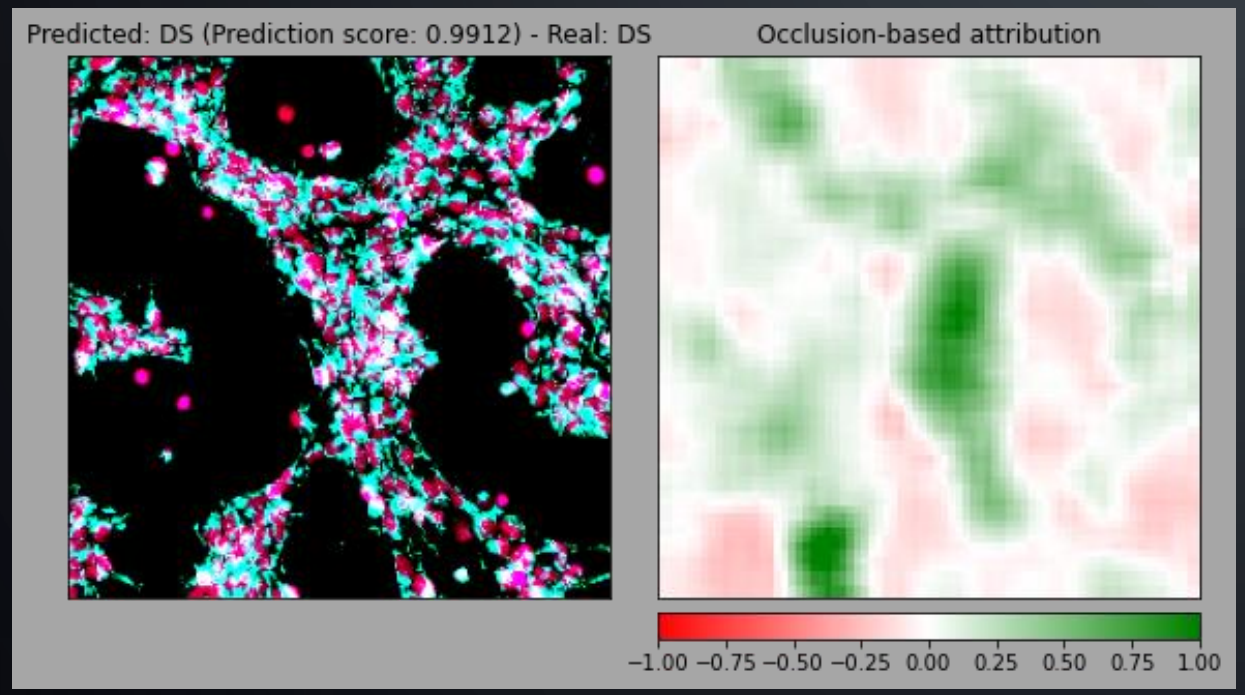
OKKLUSIONSBASIERTE ATTRIBUTION

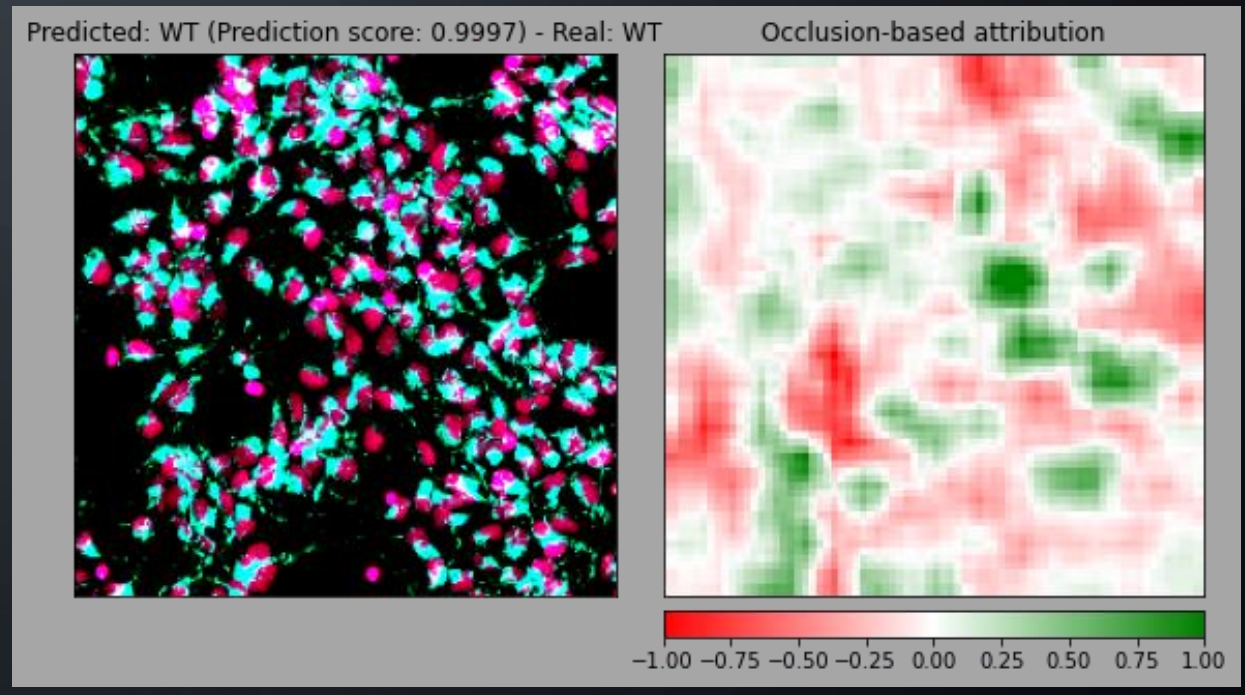
Zug wird dank der Schienen als solcher erkannt

Auto wird als Zug erkannt

OKKLUSIONSBASIERTE ATTRIBUTION

Die Strukturen der Zellkulturen werden erkannt





Problematisch: Leere Flächen

Unterschiede nicht immer intuitiv erkennbar

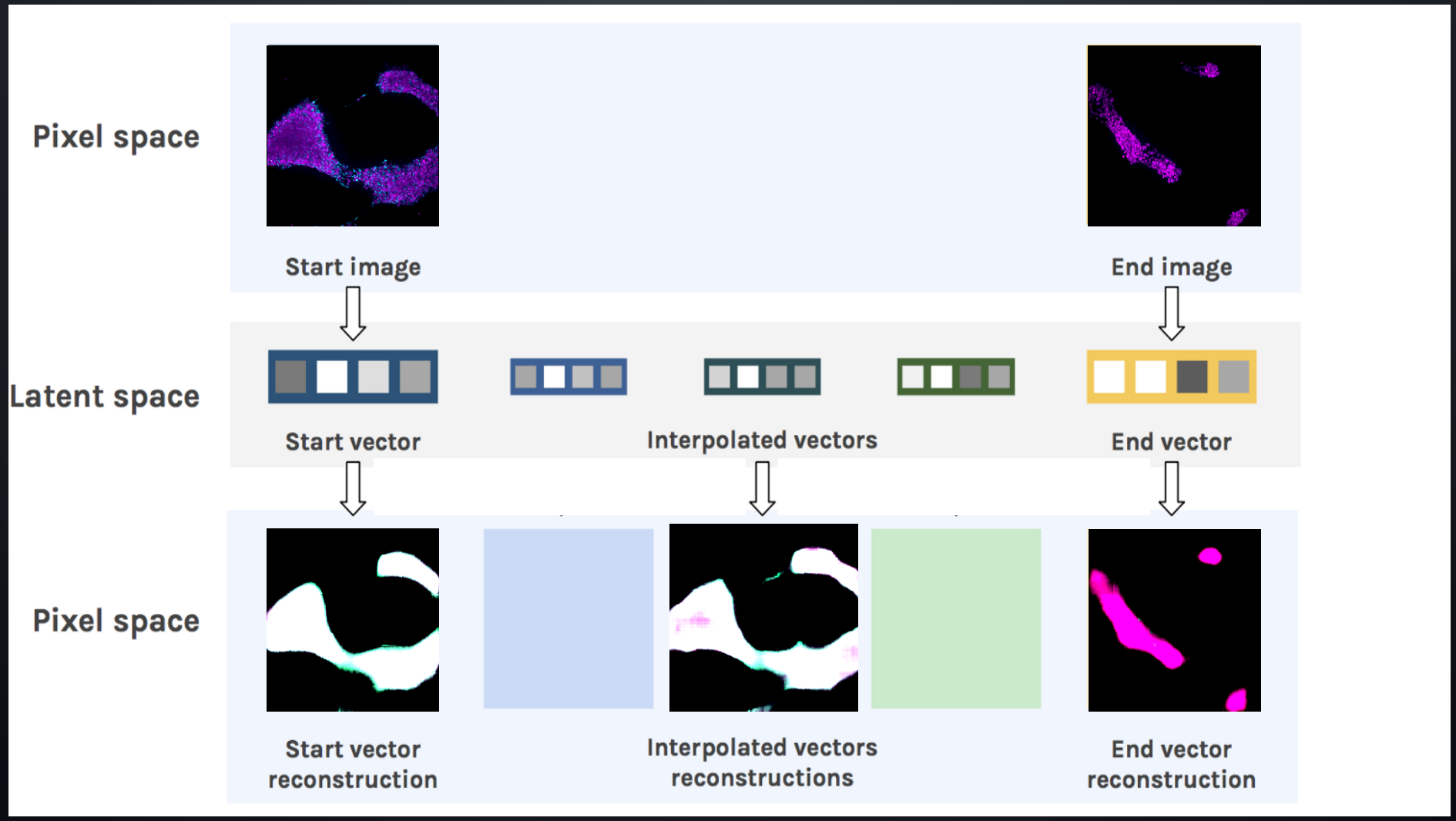
NEUER PLAN

> Eingabebild so modifizieren, dass es als andere Klasse erkannt wird

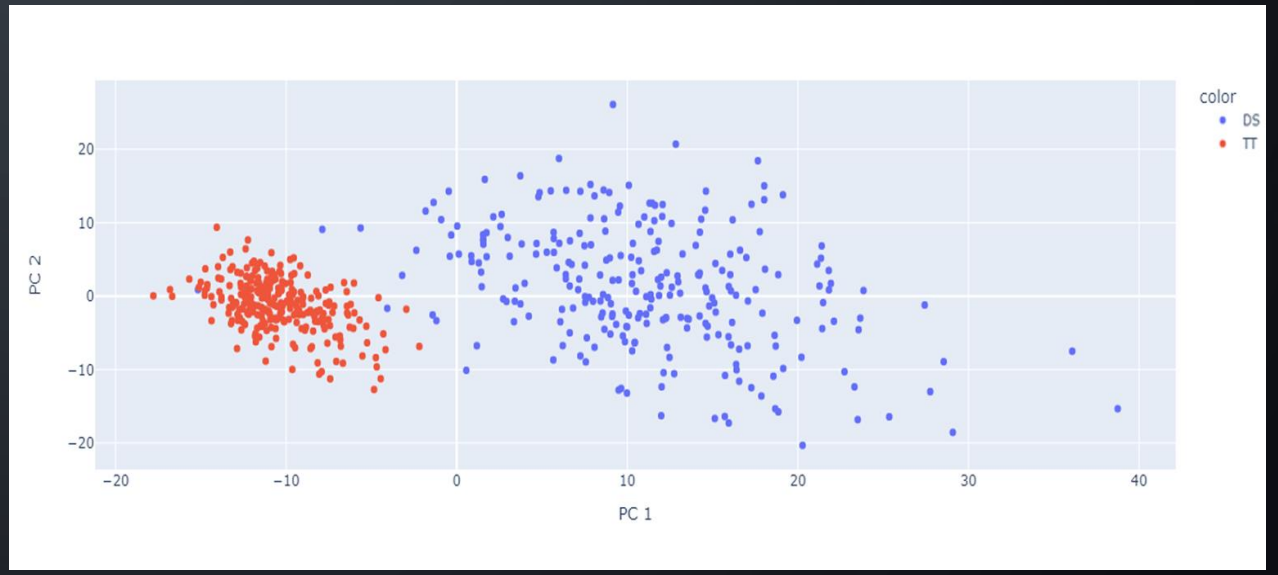
> Wie muss es sich ändern? Können wir einen Unterschied sehen?

> Neue Netzwerkarchitektur für Klassifikation & Bildgenerierung

BILDGENERIERUNG (AUTOENCODER)

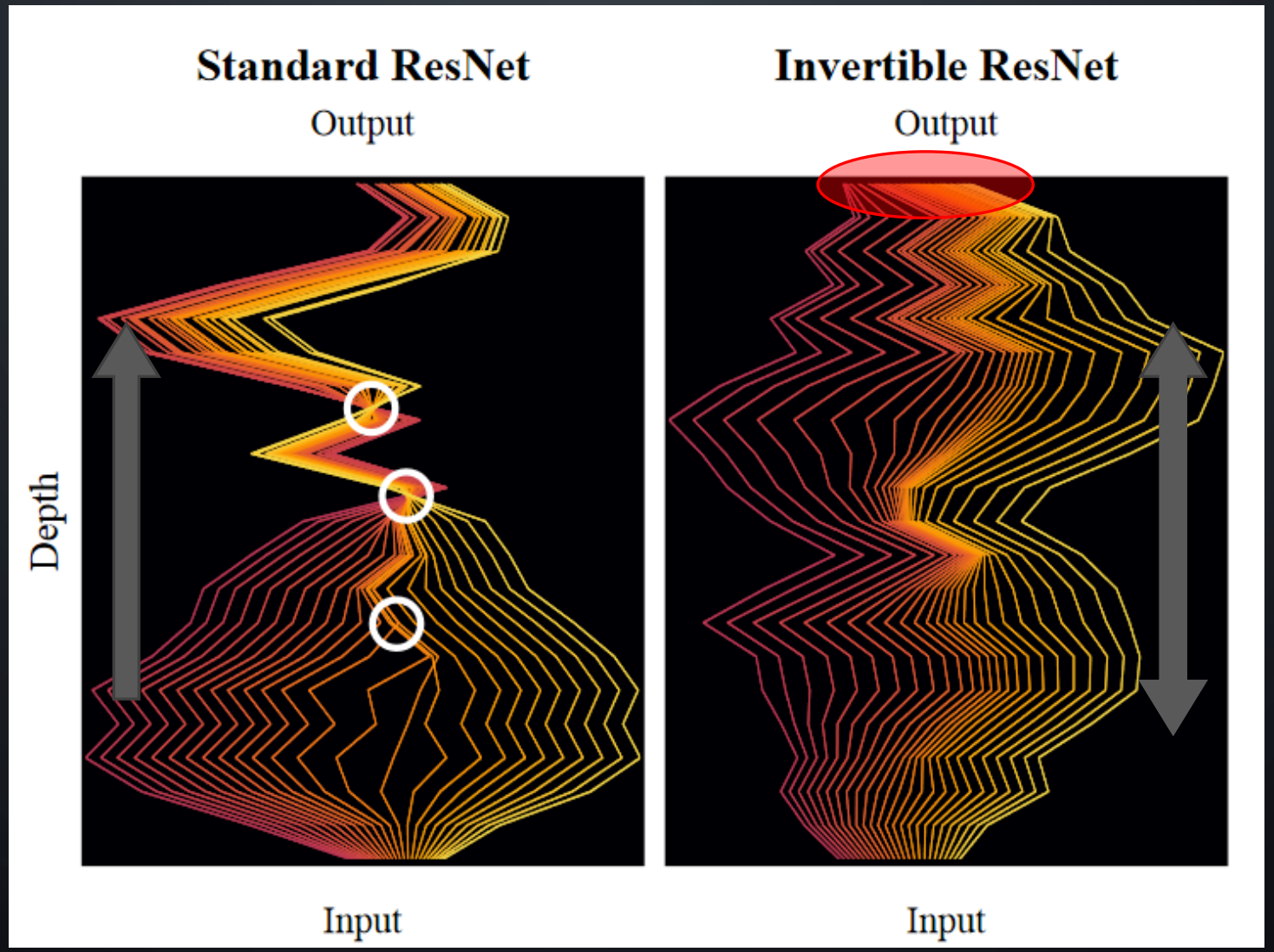


BILDGENERIERUNG (AUTOENCODER)



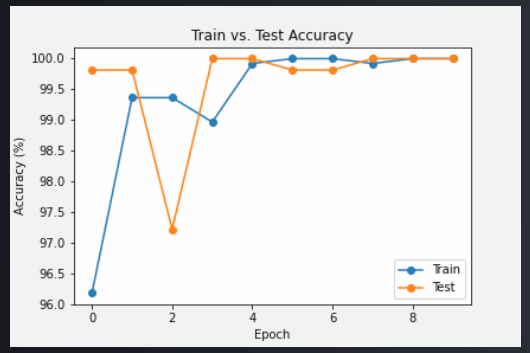
BILDGENERIERUNG (AUTOENCODER)

INVERTIBLE NEURAL NETWORK



Jens Behrmann et. all („Invertible Residual Networks“, Mai 2019)

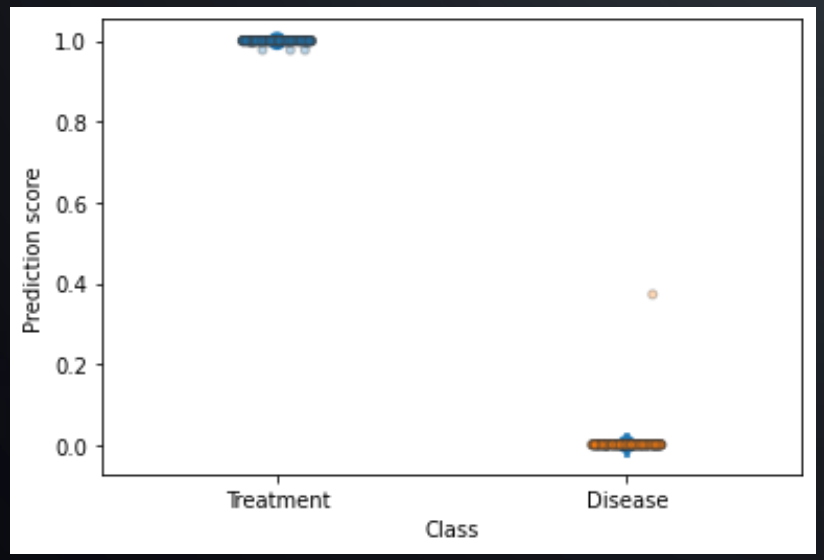
KLASSIFIKATION (DS VS. TT)

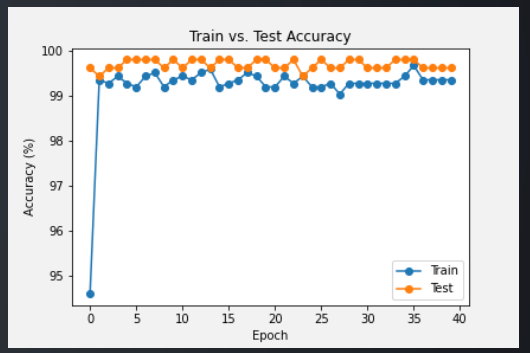


ResNext50:

540 von 540 korrekt klassifiziert

Z'-Factor: 0.924

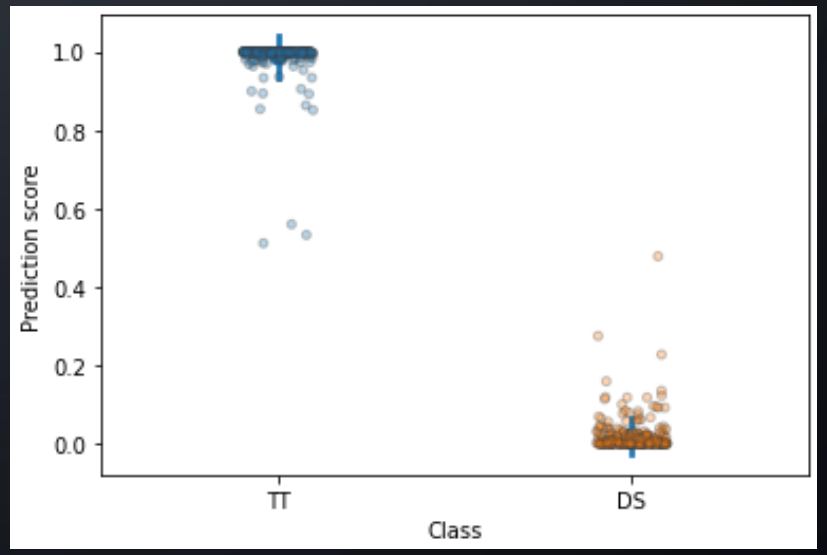




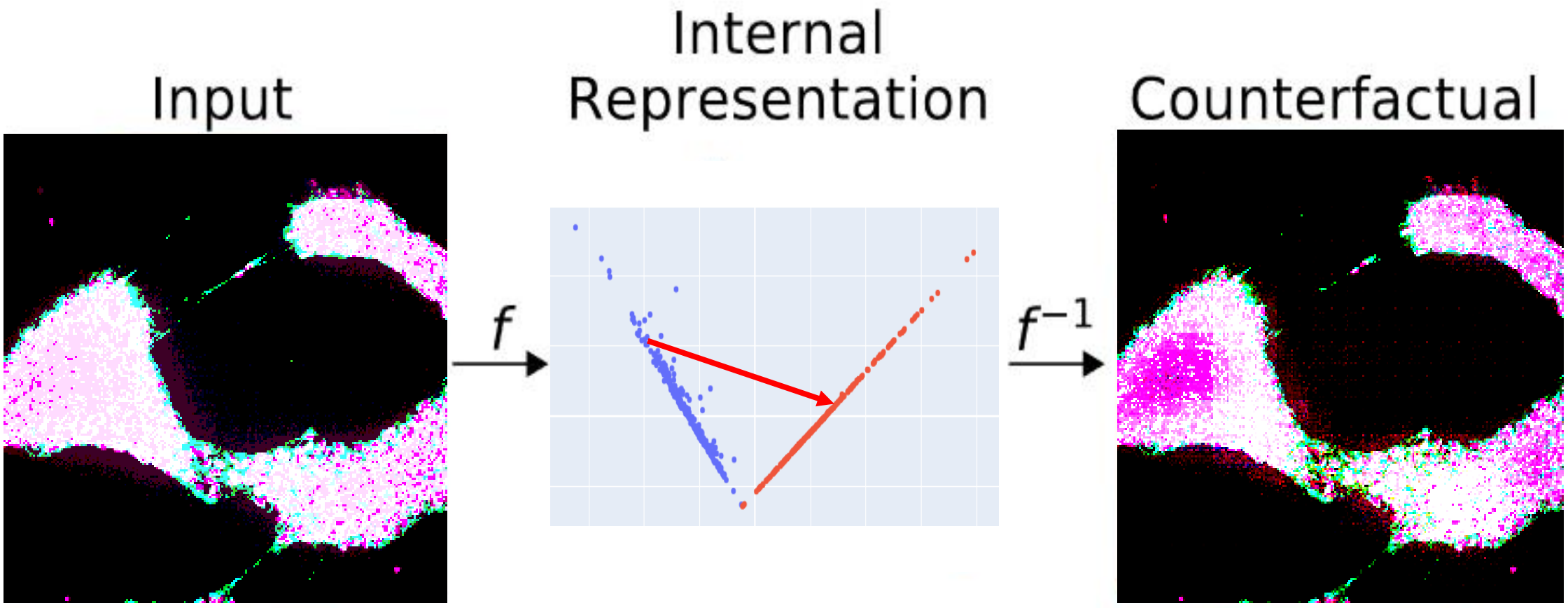
i-ResNet50:

540 von 540 korrekt klassifiziert

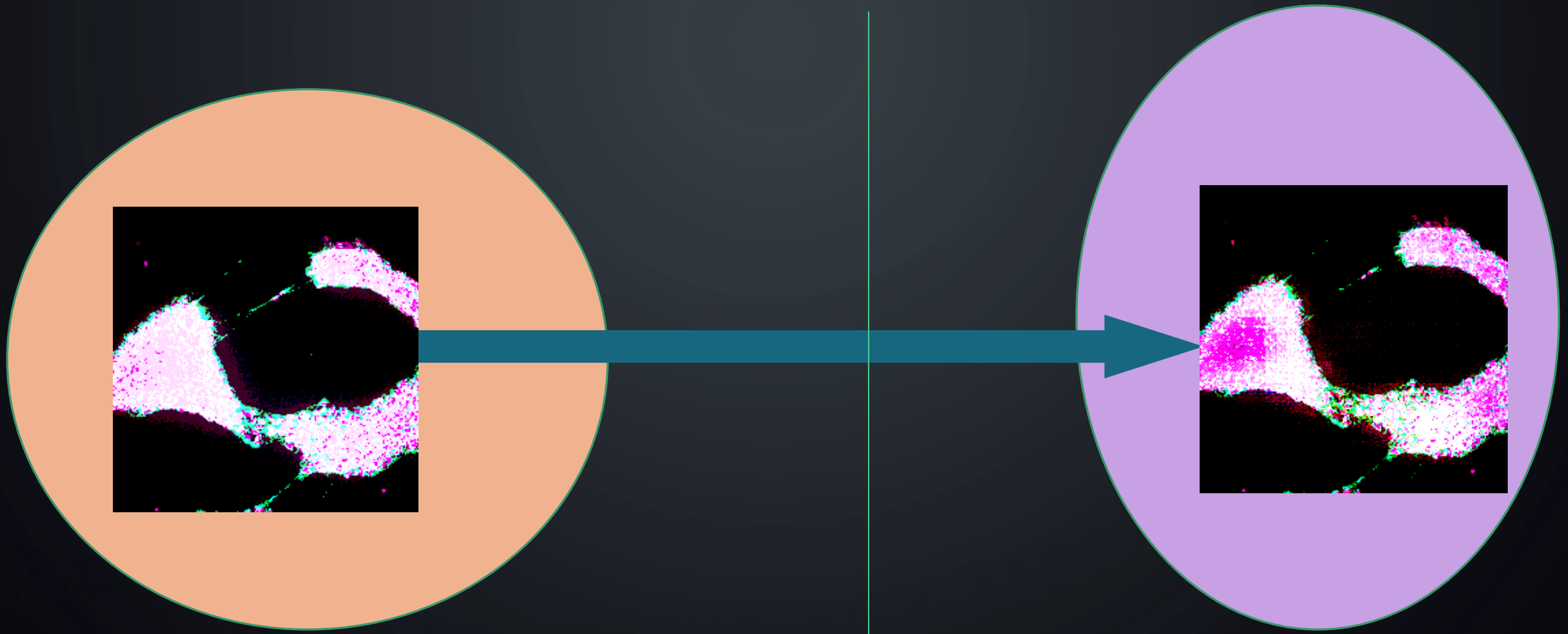
Z'-Factor: 0.701



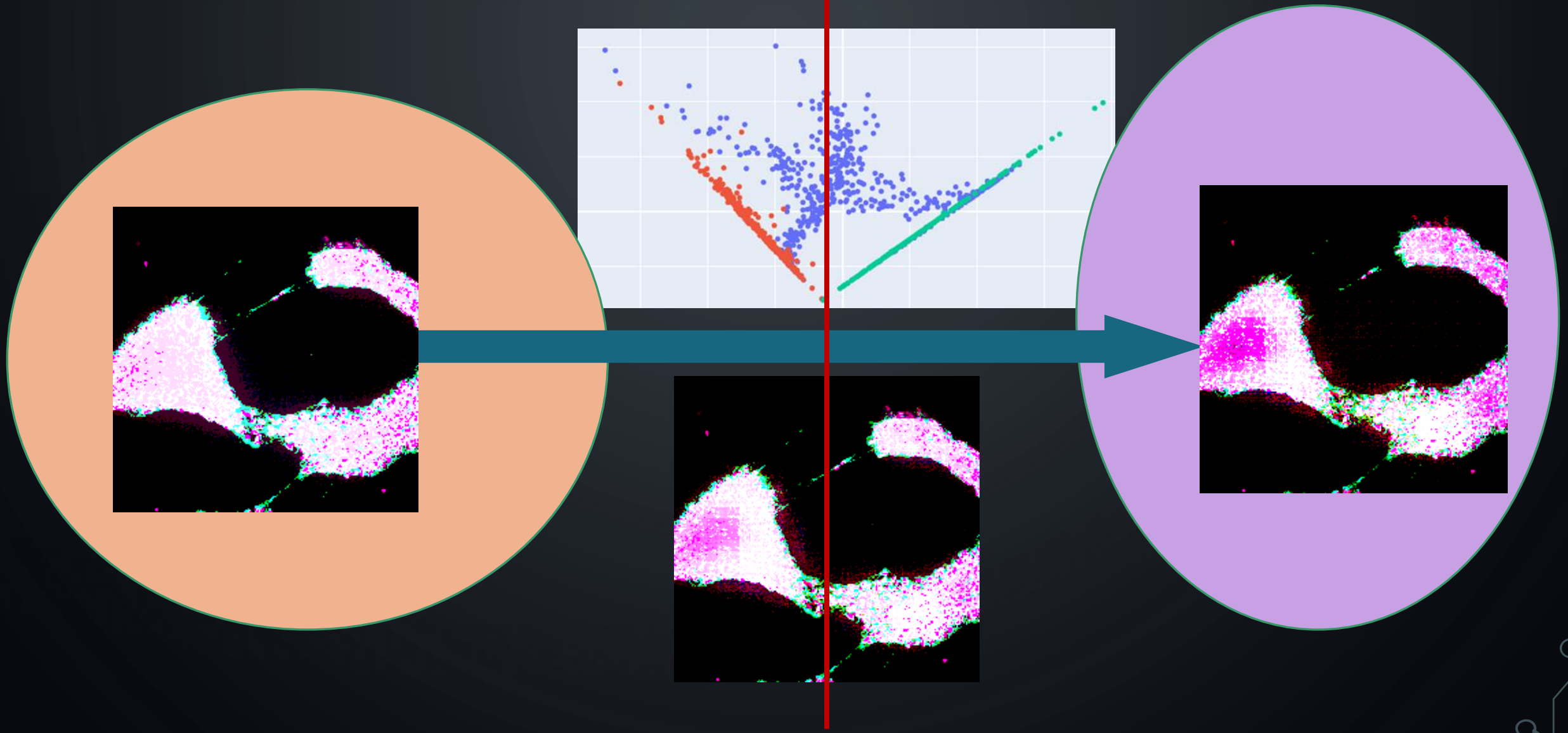
COUNTERFACTUALS



ENTSCHEIDUNGSGRENZE



ENTSCHEIDUNGSGRENZE



ZUSAMMENFASSUNG

> Eingabebild so modifizieren, dass es als andere Klasse erkannt wird

  -> Überzeugende Bilder generiert

> Wie muss es sich ändern? Können wir einen Unterschied sehen?

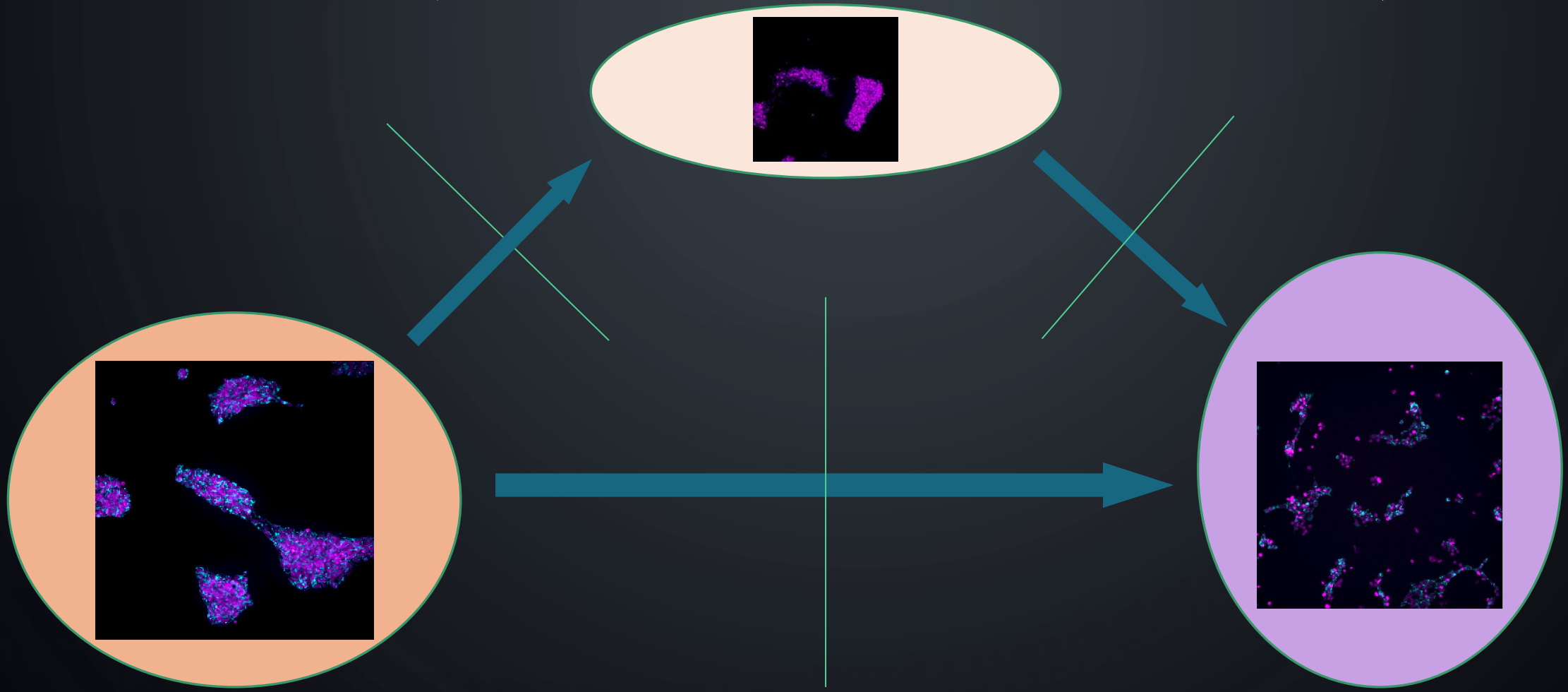
  -> Höhere mitochondriale Membranaktivität in Zellclustern bei Behandlung

> Neue Netzwerkarchitektur für Klassifikation & Bildgenerierung

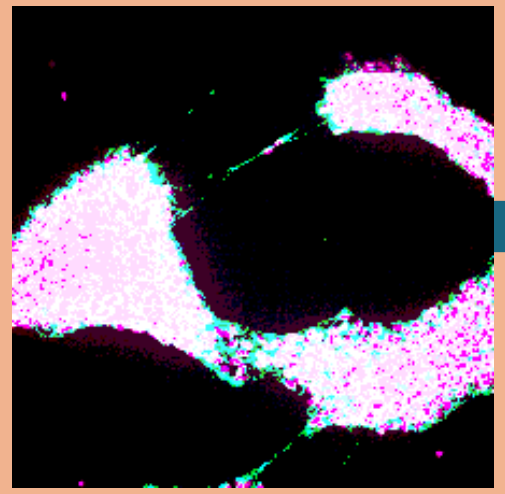
  -> Klassifikation: Vergleichbare Ergebnisse

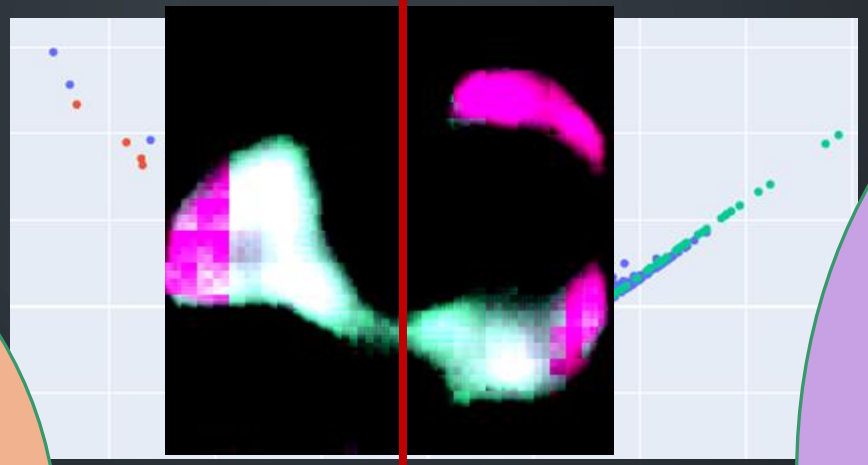
  -> Bildgenerierung: MSE-Verlust von <0.002

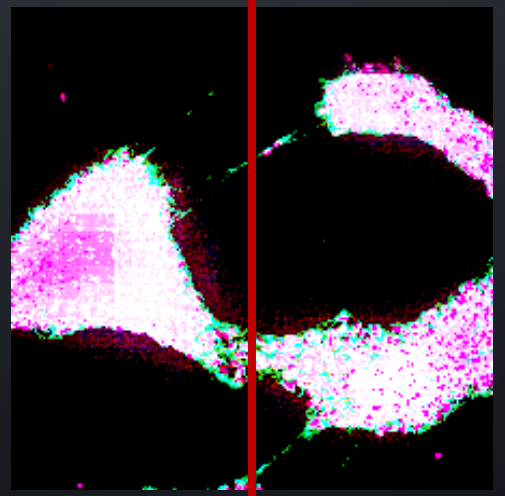
AUSBLICK: MULTI-CLASS-PROBLEM (VERSCHIEDENE MEDIKAMENTE TESTEN)

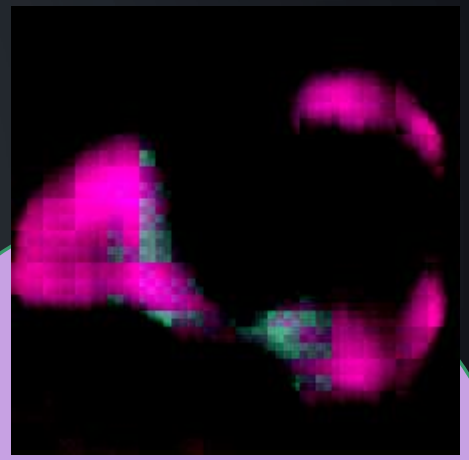


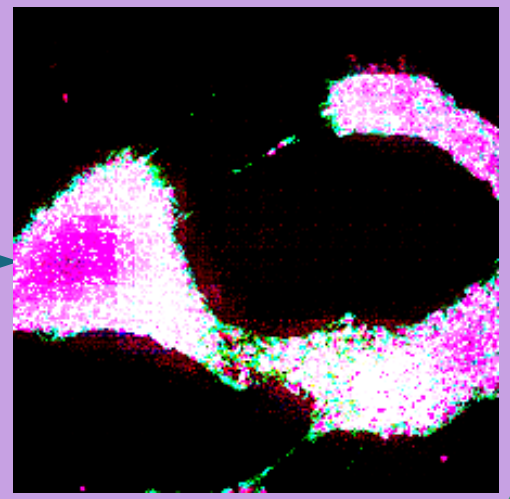
ENTSCHEIDUNGSGRENZE





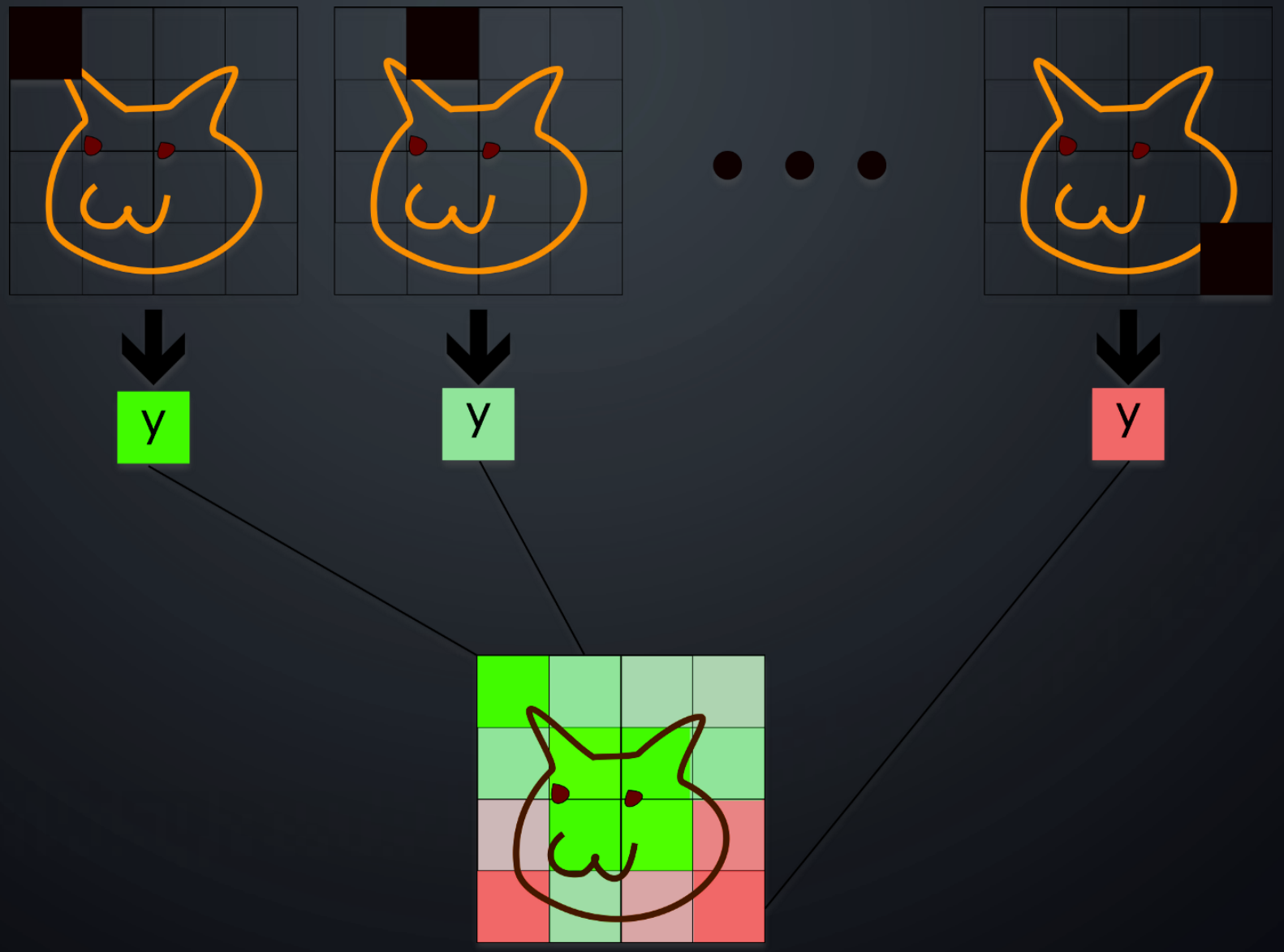




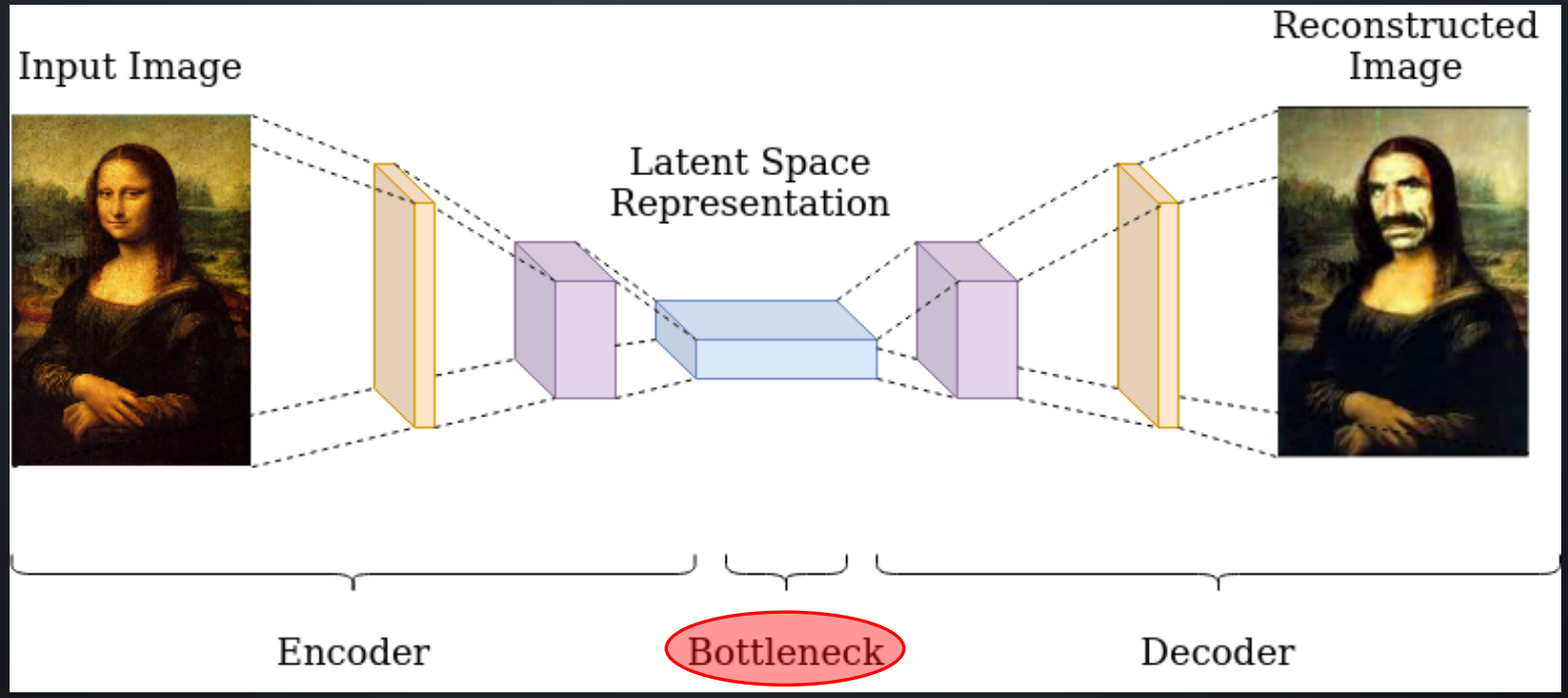


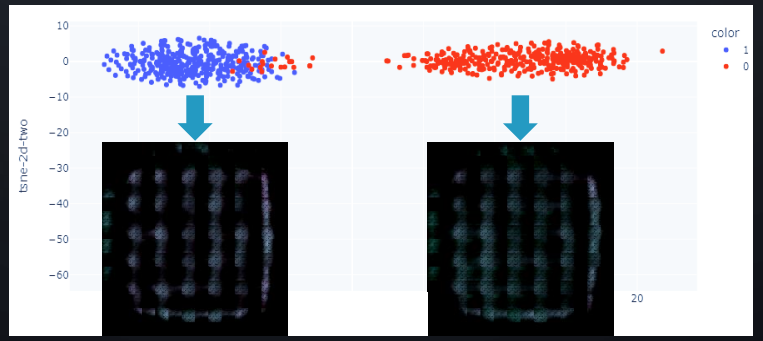
ATTRIBUTIONS

Occlusion Example:

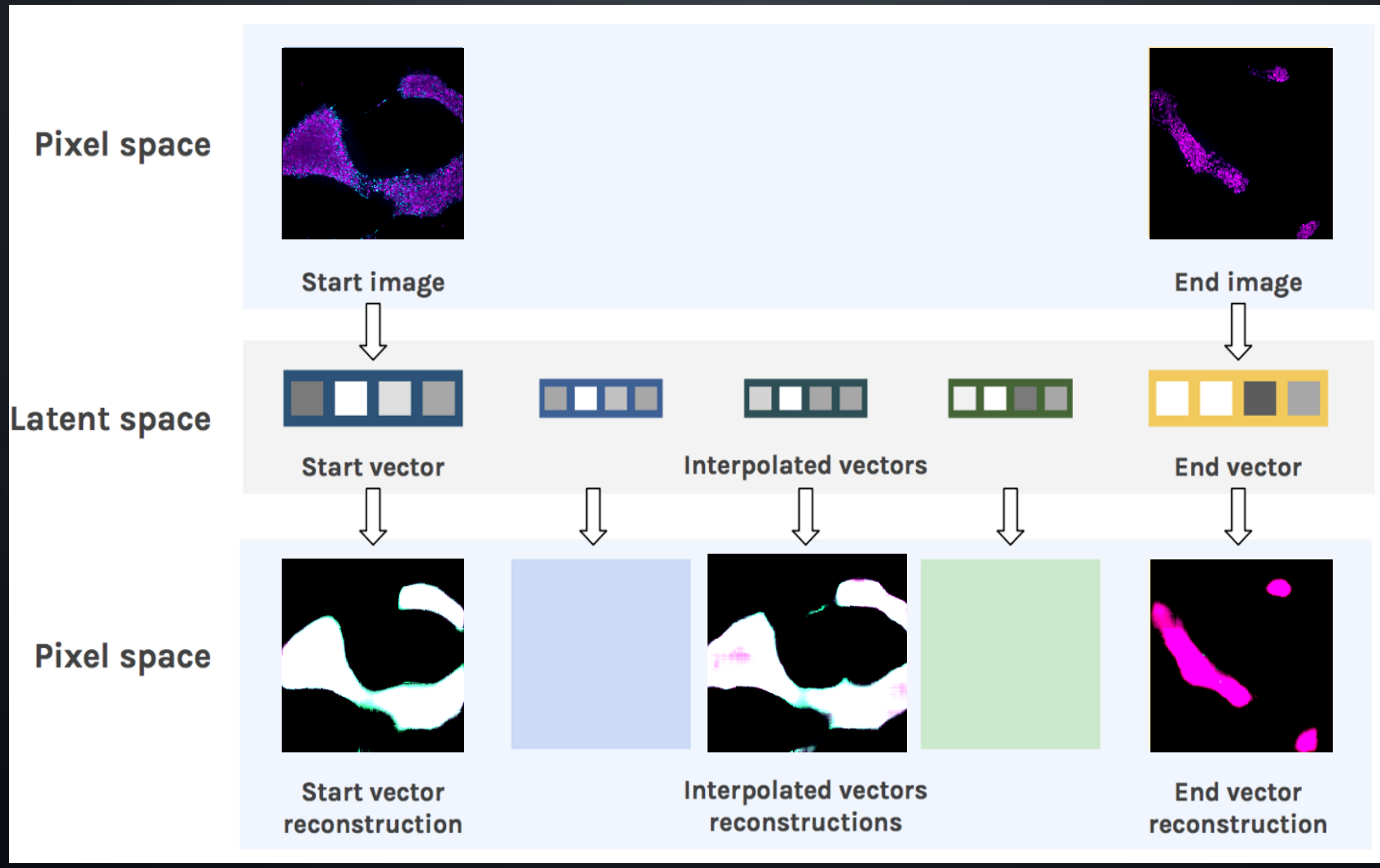


AUTOENCODER





GENERATE NEW IMAGES (AUTOENCODER)



COUNTERFACTUALS

Disease FCCP-Treated