UNIVERSITÀ POLITECNICA DELLE MARCHE

FACOLTÀ DI INGEGNERIA

Corso di Laurea triennale in *Ingegneria Informatica e dell'Automazione*

Sviluppo di un modello GAN per la generazione di immagini e relativa segmentazione

Develop of a GAN model for image generation with related segmentations

Relatore:
Prof. Adriano Mancini

Laureando:
Massimiliano Biancucci

Anno Accademico 2022/2023

Prefazione

Il mio percorso nel campo dell'inteligenza artificiale è iniziato diversi anni fà, alle superiori per l'esattezza, dove sentii per la prima volta parlare di reti neurali, ad un corso pomeridiano voluto dal prof. Roberto Lulli il quale mi ha mostrato per primo questo affascinante campo di ricerca.

Ho svolto durante il mio percorso di studi diversi progetti incentrati su questa tematica, partendo da semplici reti neurali, e confrontandomi con progetti sempre più complessi fino ad arrivare ai modelli generativi basati sull'architettura GAN (Generative Adversarial Network), del quale in questa tesi proporò una variante.

Lo scopo di questa tesi è quello di investigare la fattibilità di una potenziale soluzione ad uno dei grandi problemi che affligge oggi le aziende che si occupano di addestrare modelli neurali per la segmentazione di immagini, ovvero la difficoltà nel reperire immagini annotate, le quali hanno elevatissimi costi di realizzazione.

Tale scelta è stata naturale, in quanto ho dovuto confrontarmi in prima persona con questo problema nell'ultimo anno, come sviluppatore presso l'azienda Cloe.ai. In questa esperienza ho gestito per quasi un anno la realizzazione di un complesso dataset per l'addestramento di un modello di segmentazione di difetti, tale dataset aveva dei requisiti molto alti, e al contempo le risorse per realizzarlo erano limitate, per tale ragione ho potuto comprendere affondo le problematiche legate a questo tipo di progetto, facendo i conti io stesso con le spese e i progressi ottenuti.

La realizzazione di un dataset su larga scala è un'operazione molto complessa, che richiede una elevata coordinazione tra annotatori, revisori, sviluppatori, e un'accurata documentazione che in base al problema può richiedere anche diversi mesi per poter essere redatta efficaciemente. Tutto cio mi ha fornito la motivazione per cercare una soluzione per accorciare questo lungo e tedioso processo e dunque attenuare gli ingenti costi che un'azienda deve sostenere per realizzare un dataset di questo tipo.

Indice

Elenco delle figure

Capitolo 1

Introduzione

1.1 Il problema

Oggi nello scenario industriale e di ricerca, per l'addestramento di modelli neurali una delle principali difficoltà è la mancanza di dataset adatti al proprio task, i quali per la segmentazione di immagini possono essere particlarmente difficili da reperire o realizzare, in quanto richiedono un elevato numero di immagini annotate. Più il task è complesso più il dataset necessario per addestrare un modello sarà grande, e potenzialmente rendendo ogni singolo esempio più difficile da realizzare, richiedendo una quantità di ore uomo molto elevata.

Uno dei task in cui oggi i modelli neurali trovano un vasto impiego è la segmentazione dei difetti, principalmente in campo industriale, i quali sono utilizzati per verificare automaticamente la qualità di un prodotto o di un semilavorato, per ridurre i costi e i tempi del controllo manuale, il quale è spesso prono ad errori in quanto l'uomo è soggetto a stanchezza o a mancanza di concentrazione, al contrario un modello con un'elevata accuratezza è in grado di riconoscere un difetto con prestazioni costanti.

La problematica di avere un modello con elevata accuratezza per task di questo tipo è relativa alla preparazione dei dati. Molti task richiedono molta concentrazione da parte dell'annotatore in quanto non sempre i difetti sono ben visibili e dunque ogni singolo esempio deve essere controllato attentamente anche da più persone, e questo rende il processo di annotazione molto lungo e costoso. Spesso in oltre anche dopo diversi controlli alcuni difetti possono comunque sfuggire al processo di annotazione.

1.2 L'idea

L'obbiettivo di questo progetto è proporre una soluzione a questo problema, o almeno mitigare la sua gravità. Per far ciò si propone di realizzare un architettura in grado di generare un oggetto in un'immagine presistente, specificando la maschera di quest'ultimo.

Tale architettura deve essere in grado di prendere in input un'immagine e una maschera, e generare l'oggetto all'interno della maschera, lasciando il più possibile invariata l'immagine al di

fuori dell'area d'interesse. Il modello dovrebbe essere in grado di generare un oggetto all'interno dell'area obbiettivo seguendo la distribuzione morfologica degli oggetti presenti nel dataset di addestramento.

Tale tecnica è principalmente pensata per tutti quei casi in cui non è semplice o possibile realizzare sinteticamente un oggetto a partire da metodi deterministici, ad esempio per la segmentazione dei difetti, i quali non possono essere copiati e incollati in una nuova mmagine senza che sia presente un evidente linea di separazione tra il difetto e il resto dell'immagine, e quindi non possono essere generati con una tecnica di tipo copy paste, senza introdurre artefatti, che potrebbero portare ad un peggioramento delle prestazioni del modello, in quanto questo associerebbe il difetto alla linea di separazione.

In questo progetto il dataset utilizzato è Severstal steel defect detection, il quale presenta campioni di acciaio più tosto simili, ma in generale questa tecnica potrebbe essere utilizzata anche per casi più complessi come il rilevmento di danni su veicoli o edifici, dove la morfologia del difetto è comune, ma la morfologia, il colore e la texture dell'oggetto su cui deve essere applicato sono differenti dal dataset di addestramento, in quest'ultimo scenario infatti non è possibile utilizzare una tecnica di tipo copy paste, in quanto il difetto non potrebbe essere applicato in modo coerente con l'immagine base.

Capitolo 2

Stato dell'arte

Capitolo 3

Strumenti e metodi

3.1 Il Dataset: Severstal steel defect detection

L'acciaio è uno dei materiali più comunemente utilizzati in tutto il mondo, e la sua produzione è in continua crescita. La sua versatilità e la sua resistenza lo rendono un materiale molto utilizzato in diversi settori, come l'edilizia, l'automotive, l'industria elettronica, l'industria aerospaziale, ecc. Per produzioni su larga scala di acciaio come di altri materiali o prodotti, è necessario che il materiale sia di qualità, e che non contenga difetti, ma è difficile per gli operatori umani rilevare difetti come graffi, crepe, ecc. con elevata affidabiltà, per tale ragione è necessario adottare sistemi automatizzati che siano in grado di rilevare difetti in modo affidabile e veloce. Severstal è una delle principali aziende produttrici di acciaio in Russia, e produce circa 10 milioni di tonnellate di acciaio all'anno, quest'ultima ha iniziato dunque ad utilizzare il machine learning per automatizzare il processo di rilevazione dei difetti, e ha messo a disposizione questo dataset nel 2019 per permettere a chiunque di partecipare alla challenge, e di testare le proprie idee, sperando di riuscire ad aumentare l'affidabiltà del proprio sistema di rilevazione difetti.

Questo dataset è diviso in 2 parti, training e test set, ma per il test set non sono stati rilasciati i ground truth, quindi non è possibile testare il modello finale sul test set originale e verranno dunque utilizzati solo i dati del training set, i quali verranno suddivisi in training e test set. Il training set originale è composto da 12568 immagini, che verrano divise in 50% per il nuovo training set e 50% per il nuovo test set, quindi un totale di 6284 immagini per il training set e 6284 immagini per il test set.

il training set verrà utilizzato per il training del generatore di difetti sintetici, mentre il test set verrà utilizzato per testare il modello di segmentazione finale addestrato con il dataset sintetico.

3.2 Lama:

3.3 Stylegan2:

Capitolo 4

Sviluppo del progetto

Capitolo 5

Risultati

Bibliografia

[1] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein gan, 2017.

[2] Rinon Gal, Dana Cohen, Amit Bermano, and Daniel Cohen-Or. Swagan: A style-based wavelet-driven generative model, 2021.

[3] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks, 2014.

[4] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium, 2018.

[5] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks, 2018.

[6] Lars Mescheder, Andreas Geiger, and Sebastian Nowozin. Which training methods for gans do actually converge?, 2018.

[7] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks, 2016.

[8] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans, 2016.

[9] Roman Suvorov, Elizaveta Logacheva, Anton Mashikhin, Anastasia Remizova, Arsenii Ashukha, Aleksei Silvestrov, Naejin Kong, Harshith Goka, Kiwoong Park, and Victor Lempitsky. Resolution-robust large mask inpainting with fourier convolutions, 2021.

[10] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric, 2018.