



**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'intelligenza artificiale è iniziato diversi anni fa, 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 efficacemente. Tutto ciò 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

<b>Prefazione</b>	<b>ii</b>
<b>Indice</b>	<b>iii</b>
<b>1 Introduzione</b>	<b>1</b>
1.1 Motivazione . . . . .	1
1.1.1 Modelli neurali per la segmentazione nel controllo qualità . . . . .	1
1.1.2 Il dataset, requisiti e problematiche di realizzazione . . . . .	2
1.1.3 Approccio al problema . . . . .	3
1.2 Introduzione alle reti neurali . . . . .	6
1.2.1 Dal machine learning al deep learning . . . . .	6
1.2.2 Confronto con le reti neurali biologiche . . . . .	7
1.2.3 Feedforward neural networks . . . . .	8
1.2.4 Convolutional neural networks . . . . .	9
<b>2 Stato dell'arte</b>	<b>10</b>
<b>3 Strumenti e metodi</b>	<b>11</b>
3.1 Il Dataset: Severstal steel defect detection . . . . .	11
3.2 Lama: . . . . .	11
3.3 Stylegan2: . . . . .	11
<b>4 Sviluppo del progetto</b>	<b>12</b>
<b>5 Risultati</b>	<b>13</b>
<b>Elenco delle figure</b>	<b>14</b>

# Capitolo 1

## Introduzione

### 1.1 Motivazione

#### 1.1.1 Modelli neurali per la segmentazione nel controllo qualità

Oggi le reti neurali trovano un vasto impiego in moltissimi campi, dall'industria alla medicina, fino alla vita di tutti i giorni. Il grande vantaggio che ci portano è la capacità di apprendere da un set di dati, e di generalizzare su di uno nuovo, permettendoci di risolvere problemi che altrimenti sarebbero matematicamente troppo complessi da risolvere con un algoritmo. Ci sono vari esempi in cui i modelli neurali raggiungono risultati superiori a quelli ottenuti dall'uomo, in determinati task, o almeno se non lo superano in termini di accuratezza, lo fanno in termini di velocità, scalabilità, costi e prestazioni.

Un task in cui le reti neurali eccellono è la segmentazione di immagini, ovvero la classificazione pixel per pixel di un'immagine, questo tipo di task è utilizzato ad esempio nel campo medico per la segmentazione di organi, tumori, o in campo industriale per la segmentazione di difetti, per la verifica automatica della qualità di un prodotto o di un semilavorato.

Nel caso specifico, per la segmentazione dei difetti l'utilizzo di questo tipo di modelli è molto diffuso, in quanto risolve un grave problema che affligge i reparti controllo qualità delle aziende, ovvero il calo della concentrazione al quale un operatore è soggetto dopo un certo numero di ore di lavoro. Infatti una persona per quanto allenata e preparata, dopo un certo numero di ore di lavoro, è soggetta a stanchezza e con essa la sua accuratezza nel riconoscere un difetto diminuisce, mentre un modello neurale adeguatamente addestrato, in condizioni ambientali stabili, come ad esempio una adeguata illuminazione, una videocamera ad alta risoluzione e un'adeguata distanza dal soggetto, sarà in grado di mantenere un'accuratezza costante, senza necessità di fermarsi per riposare. Questo si traduce in un risparmio di tempo e di denaro per l'azienda, In quanto il controllo manuale richiede più tempo ed è più soggetto ad errori, i quali spesso si trasformano in ritardi nella consegna dei prodotti, spese di trasporto aggiuntive per il ritorno o la sostituzione del prodotto, o addirittura la perdita di un cliente.

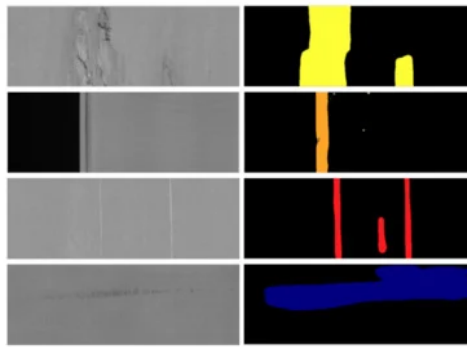


Figura 1.1: An example of segmented defects from the "Severstal steel defect dataset". credits: Neven Robby and Goedemé Toon, 2021, A Multi-Branch U-Net for Steel Surface Defect Type and Severity Segmentation. <https://www.mdpi.com/2075-4701/11/6/870>

### 1.1.2 Il dataset, requisiti e problematiche di realizzazione

La problematica di avere un modello con elevata accuratezza per task di segmentazione è relativa alla quantità e qualità dei dati necessari, i quali raramente sono disponibili opensource o per l'acquisto, rendendo necessaria la creazione di un dataset apposito. Molti task richiedono un grande quantità di dati per essere generalizzati correttamente, e ogni singolo esempio richiede molta concentrazione da parte dell'annotatore in quanto non sempre i difetti sono ben visibili.

In generale la creazione di un dataset è un'operazione molto complessa, non solo dal punto di vista dell'annotatore, ma in oltre da un punto di vista organizzativo e logistico. Infatti ci sono diversi step che si devono seguire:

- **Acquisizione delle immagini:** Le immagini devono essere acquisite in modo da avere una buona qualità, o almeno sufficiente ai fini dell'apprendimento. Se possibile in oltre dovrebbero avere una adeguata uniformità di condizioni (luce, distanza, ...) per garantire le migliori prestazioni da parte del modello, ovviamente solo se poi è possibile garantire le stesse condizioni anche nell'utilizzo finale del modello, altrimenti una grande varietà delle condizioni è preferibile.
- **Definizione delle classi:** Nel caso di dataset multi-classe, uno step molto importante è quello di scegliere accuratamente le classi e definire in maniera univoca l'associazione tra una classe e una particolare tipologia di difetto. Questo passaggio potrebbe sembrare banale ma in realtà nasconde delle grandi insidie, infatti una classificazione non adeguata andrà a causare confusione nel modello, diminuendo la sua accuratezza e/o rendendo il lavoro più difficile per gli annotatori andando a rallentare il processo di annotazione o comunque a ridurne la qualità. Questo tipo di problematiche purtroppo si manifestano chiaramente soltanto in uno stato avanzato del progetto, rendendo necessarie revisioni della documentazione, modifica di tutti gli esempi già annotati, con conseguente perdita di tempo e denaro.
- **Definizione della documentazione:** Questo passaggio è un'estensione del precedente, e consiste nella definizione di una documentazione che specifichi senza ambiguità, ad un nuovo annotatore come riconoscere senza dubbio un difetto e classificarlo nella giusta

classe. Questa fase spesso non termina prima dell'inizio dell'annotazione, ma si protrae per tutta la durata del progetto, in quanto spesso nuovi casi non previsti si presentano durante l'annotazione, e la documentazione deve essere aggiornata in tempo reale.

- **Annotazione:** Questo è il passaggio più lungo e costoso, in quanto richiede una squadra di persone, che devono essere formate per lo specifico task, e che devono essere costantemente seguite per garantire la qualità del lavoro.
- **Revisione:** Assieme all'annotazione questo è un passaggio chiave, in quanto permette di verificare che l'annotazione sia stata fatta correttamente, e che non ci siano errori nell'annotazione. Spesso infatti gli annotatori acquisiscono dei bias errati nei confronti di una certa classe, o di un certo tipo di difetto, che deve essere identificato e reso noto all'annotatore per correggerlo, ed evitare che questo errore si ripeta in futuro. Per evitare che ciò accada oltre al primo annotatore lo stesso esempio viene solitamente rivisto da 2 o 4 persone diverse. Si noti che gli errori degli annotatori che non vengono identificati verranno appresi dal modello finale come una corretta classificazione, ciò giustifica un tale dispendio di risorse in questa fase.

### 1.1.3 Approccio al problema

La creazione di un dataset come precedentemente illustrato è un processo complesso e dispendioso, che richiede molte risorse umane e finanziarie, dunque l'intento in questo progetto è quello di proporre un approccio alternativo che sia in grado di ridurre per quanto possibile la durata e il costo di questo lavoro. Partendo dal presupposto che almeno in parte il dataset deve essere realizzato manualmente, la proposta è quella di realizzare una certa quantità di campioni manualmente seguendo lo schema già visto, per poi addestrare un modello neurale per generare ulteriori esempi sintetici, raggiungendo un numero di esempi totali che permetta di addestrare un modello con buone prestazioni, ad un costo ridotto rispetto al caso in cui tutti i dati fossero stati realizzati manualmente.

Per la definizione della pipeline di generazione dei dati, si è partiti dal concetto di *generative adversarial network* (GAN), che è una tecnica di machine learning che permette di generare dati sintetici utilizzando come base dati reali, tali dati sintetici possono essere utilizzati per addestrare un modello neurale. Tale tecnica ha trovato riscontri positivi in molte ricerche pubblicate in ambito di computer vision [? ], in cui i modelli GAN vengono utilizzati per espandere il numero di immagini presenti in un dataset e migliorare la generalizzazione di un modello di classificazione. Ovviamente gli aumenti di accuratezza, precisione e recall dipendono dal numero di esempi presenti nel dataset e dalla complessità del problema. Tale tecnica potrebbe essere considerata una versione più sofisticata di data augmentation, in quanto permette di generare dati sintetici molto più complessi e realistici di quelli che si possono ottenere con semplici trasformazioni geometriche o matematiche. Per generare dati utilizzabili per addestrare un modello di segmentazione però è necessario risolvere un'ulteriore problema, infatti un normale modello GAN, fedele alla sua definizione originale [? ], è in grado di generare intere immagini, che possono essere utilizzate per addestrare un modello di classificazione, ma non sono utilizzabili per addestrare un modello di segmentazione, in quanto per l'addestramento

di tale architettura è necessario che gli oggetti di interesse abbiano una maschera che specifichi la loro posizione. Ci sono vari approcci di augmentation per la segmentazione che risultano molto più semplici di addestrare un modello GAN, come il caso del metodo "copy paste" [?], il quale propone come augmentation per i dataset di segmentazione la copia di un oggetto presente in un'immagine, ritagliandolo attraverso la sua maschera, e incollandolo su di un nuovo background potenzialmente in una nuova posizione, tale metodo risulta estremamente efficace per oggetti indipendenti dal contesto con contorni ben definiti, ma risulta inutile nel momento in cui l'oggetto che vogliamo generare ha una interdipendenza forte con l'area immediatamente circostante, pensiamo ad esempio un difetto su di un'auto, un graffio o una bozza, non potrà essere copiato da un'auto e incollato su di un'altra in quanto subentreranno una serie di di artifacts, come la variazione netta di colore tra l'auto e il difetto, rischiando di introdurre un bias nel modello, il quale finirebbe per cercare la variazione netta di colore e non più le features del difetto. Per risolvere questo problema con questa particolare categoria di dataset ci sono 2 principali strade illustrate di seguito.

### **Generatore con architettura a solo decoder**

Questo approccio prevede un'architettura a solo decoder, ovvero un modello che prende in ingresso un tensore di determinate dimensioni e che attraverso una serie di operazioni di upsampling o dilated convolution ad esempio, effettua un'espansione di tale tensore portandolo alle dimensioni finali. Generalmente si mette in ingresso un vettore casuale di dimensione definita, ottenendo in uscita un tensore delle dimensioni di un'immagine con i canali rgb ed eventualmente altri n canali per la maschere che identificano le classi desiderate. Un'esempio di tale architettura è illustrata di seguito.

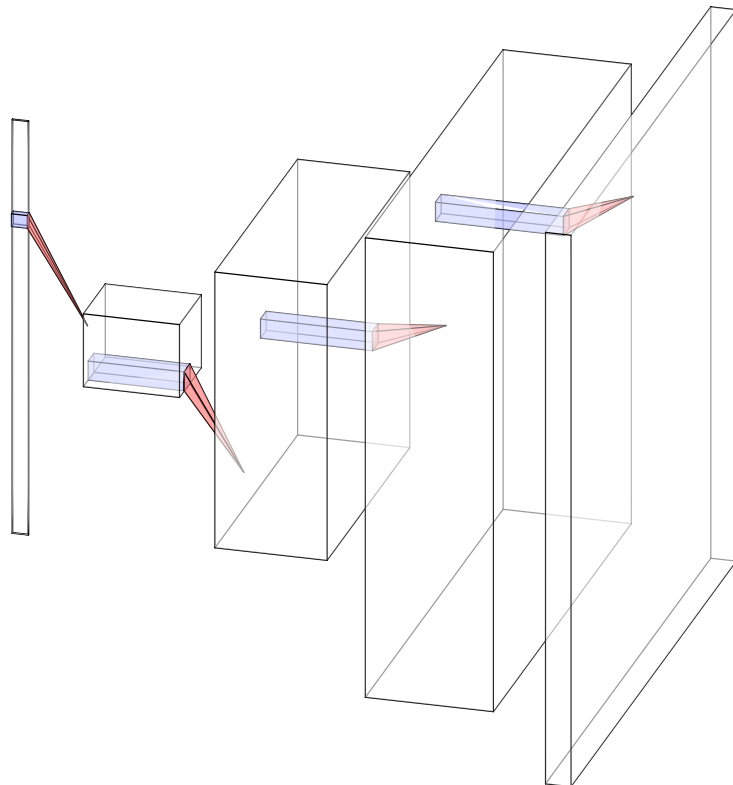


Figura 1.2: Esempio di architettura a solo decoder.



Questo approccio risulta più semplice da implementare, lasciando però al modello il compito di imparare a generare correttamente le immagini e delle maschere coerenti, compito non facile, che a seconda del task può necessitare di un elevato numero di esempi. Questa architettura dovrà imparare oltre alla struttura degli oggetti target, a posizionarli nell'immagine e a generare lo sfondo. Un'altra problematica di questo approccio è il controllo, infatti l'unico modo di interagire con tale modello è modificando il valore del vettore  $z$  dato in input, il quale permette di spostarsi nello spazio latente, al quale il modello associa diverse caratteristiche dell'immagine di output in maniera altamente non lineare, rendendo un eventuale controllo dell'output del modello molto difficile. La difficoltà di controllare il modello rende dunque difficoltoso o impossibile controllare, qualora fosse necessario, la posizione, l'intensità, la dimensione o la forma degli oggetti generati.

### Generatore con architettura a encoder-decoder

Quest'ultimo è l'approccio scelto in questo progetto, in quanto permette di avere un maggiore controllo sull'output del modello, anche se prevede una training pipeline più complessa da gestire. Al contrario del caso precedente il modello con struttura encoder-decoder permette di passare in ingresso un'immagine base e una o più maschere che identificano le aree dove determinati oggetti devono essere generati, trasformando il task di generazione puro in un task di inpainting. I vantaggi principali di questa tecnica stanno nel fatto che il modello non deve più apprendere la distribuzione degli oggetti nello spazio dell'immagine, nè deve apprendere in maniera troppo approfondita i background, ma si può focalizzare maggiormente sulla struttura degli oggetti da generare.

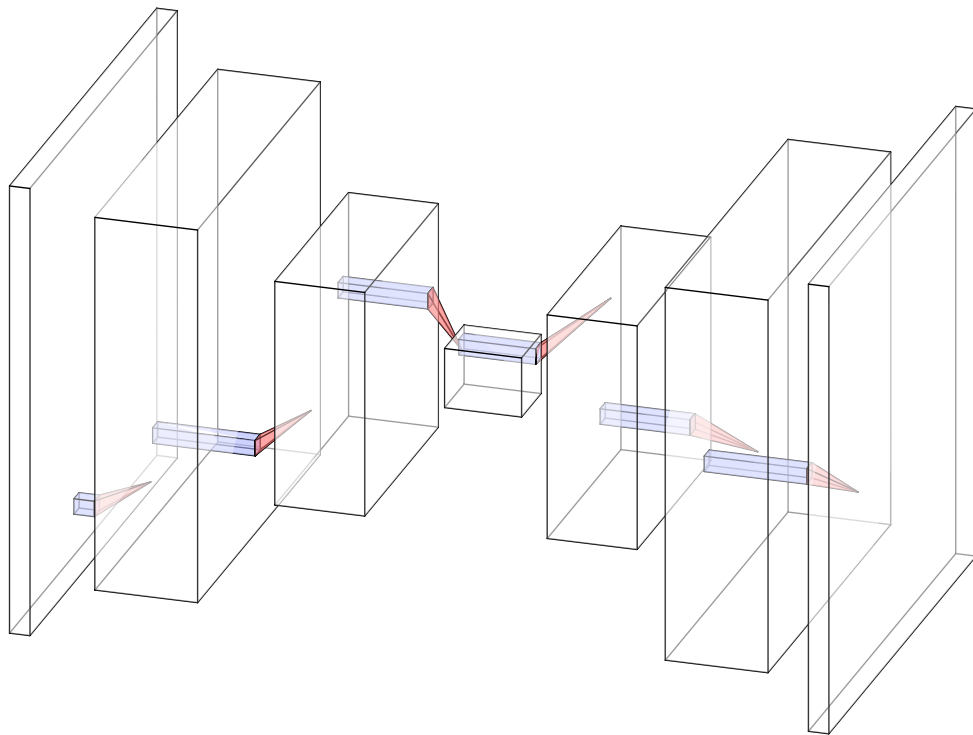


Figura 1.3: Esempio di architettura con encoder e decoder.

## 1.2 Introduzione alle reti neurali

In questa sezione verranno date le basi per comprendere cosa sono le reti neurali e come funzionano, descrivendo alcune loro varianti. Per dare al lettore senza esperienza in questo campo una panoramica generale di come funzionano questi costrutti matematici, come apprendono e come vengono utilizzate per risolvere problemi generici attraverso l'addestramento. Per prima cosa però diamo un'occhiata a dove si collocano le reti neurali nel campo dell'intelligenza artificiale.

### 1.2.1 Dal machine learning al deep learning

L'intelligenza artificiale è un campo di ricerca con l'obiettivo di risolvere una grande varietà di problemi, che per essere risolte attraverso la programmazione classica avrebbero bisogno di una grande quantità di conoscenze non disponibili, o semplicemente di troppo lavoro.

I Programmi basati sul paradigma dell'intelligenza artificiale si propongono di superare questi ostacoli acquisendo direttamente queste conoscenze dai dati grezzi, tale capacità è nota come machine learning. Sotto questa grande famiglia di algoritmi si trovano altri sottogruppi quali il representation learning e all'interno di quest'ultimo il deep learning.

Il deep learning rispetto ai metodi più classici, tipicamente in grado di riconoscere soltanto relazioni lineari (come ad esempio l'SVM o support vector machine), si propone come alternativa per l'apprendimento di funzioni non lineari anche molto complesse. Il termine "deep learning" deriva proprio dalla capacità di riuscire a cogliere queste relazioni molto "profonde" tra i dati di ingresso e uscita.

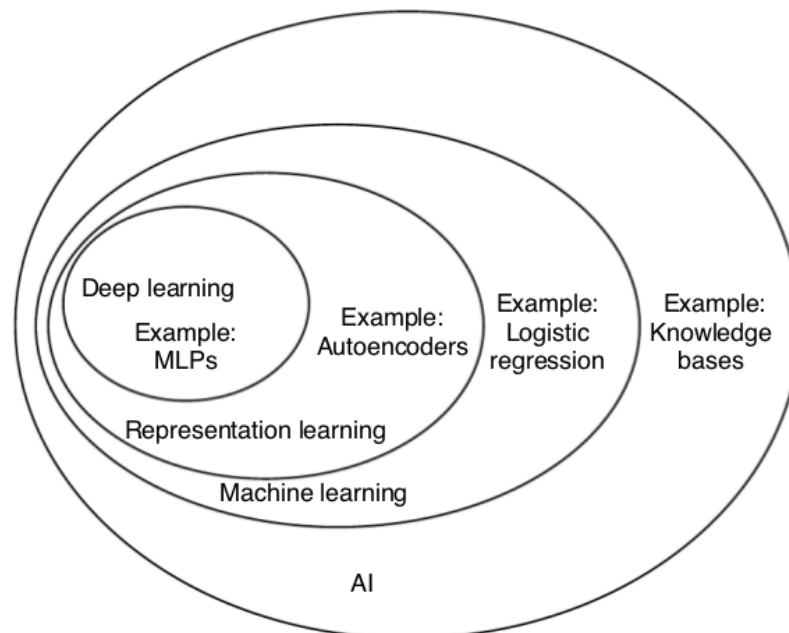


Figura 1.4: Un diagramma di ven che illustra le relazioni tra i diversi sottogruppi dell'intelligenza artificiale, vediamo infatti come il deep learning sia un sottogruppo del representation learning, che a sua volta è un sottogruppo del machine learning.  
credits: Yoshua Bengio, Ian J. Goodfellow, Aaron Courville 2015, From the book "Deep Learning"

### 1.2.2 Confronto con le reti neurali biologiche

Queste strutture matematiche fecero la prima comparsa in un articolo del 1957, pubblicato da Warren McCulloch e Walter Pitts "A logical calculus of the ideas immanent in nervous activity", articolo che gettò le basi per la costruzione di reti neurali artificiali come le conosciamo oggi partendo proprio dal sistema nervoso. Infatti le reti neurali artificiali sono ispirate alle reti neurali biologiche, che sono costituite da un insieme di neuroni, che presentano componenti simili seppur la controparte artificiale sia molto più semplice.

Il neurone biologico componente principale del cervello e del sistema nervoso, è costituito da un corpo cellulare o soma, dai dendriti, dall'assone e dalle sinapsi.

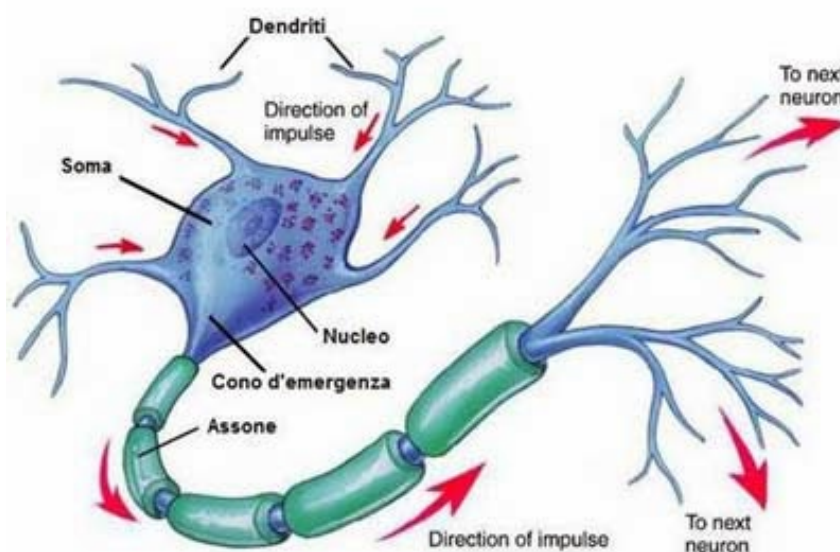


Figura 1.5: Schema di un neurone biologico.

Il neurone riceve degli impulsi da altri neuroni o da altri apparati sensoriali attraverso i dendriti, che sono delle strutture ramificate che si estendono dalla cellula, questi impulsi vengono accumulati all'interno del soma, e se la somma supera un certo valore di soglia si innesca un'azione di potenziale d'azione, che viene trasmesso attraverso l'assone verso altri neuroni o verso altri organi. L'assone è una struttura che si estende dalla cellula, a seconda della tipologia di neurone può estendersi da pochi micrometri fino anche ad un metro, e presenta all'estremità opposta del soma le sinapsi, le quali consentono la propagazione dell'impulso dall'assone ad altri neuroni.

La struttura dell'assone è rivestita dalla guaina mielinica che ne facilita la conduzione degli impulsi, maggiore è lo spessore della guaina minore è la resistenza al passaggio dell'impulso, e dunque maggiore sarà l'ampiezza del segnale in uscita a parità di quello di ingresso. Tale meccanismo è utilizzato per accumulare informazione nella struttura della rete neurale biologica.

### 1.2.3 Feedforward neural networks

Uno dei primi modelli ad essere proposti e utilizzati nella pratica è stato quello delle reti neurali feedforward (o multi layer perceptron MLP), in cui i neuroni sono disposti in strati, e l'output di ogni neurone di uno strato è connesso con l'input di tutti i neuroni dello strato successivo, attraverso delle connessioni che conservano un peso, la configurazione di tali pesi determina il comportamento della rete.

Tali reti come dice il nome propagano l'informazione dallo strato di input a quello di output attraverso i layer intermedi, in modo lineare, senza retropropagazioni intermedie dell'informazione.

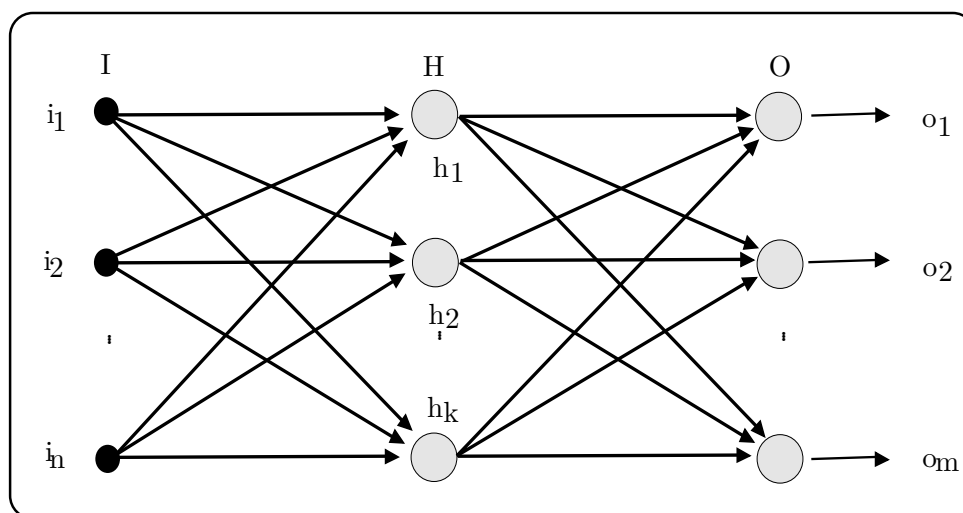


Figura 1.6: Esempio di rete neurale feedforward. In tale rete è possibile vedere i neuroni rappresentati dai nodi del grafo, e le interconnessioni tra di essi che definiscono il peso o l'importanza di tale connessione.

File:Rete-Neurale2.svg. In Wikipedia.

<https://commons.wikimedia.org/wiki/File:Rete-Neurale2.svg>

Andiamo ora a scomporre le varie componenti di tale modello e illustrarne il funzionamento, cominciando dal neurone artificiale.

## Il neurone artificiale

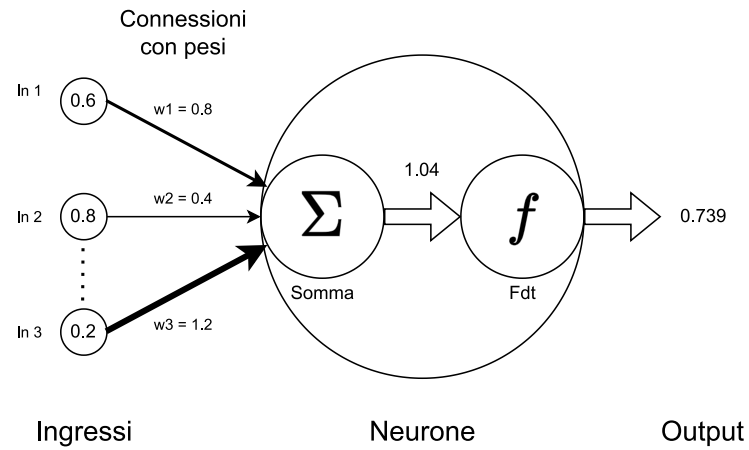


Figura 1.7: Esempio di Neurone artificiale.

### 1.2.4 Convolutional neural networks

TODO.

## 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 affidabilità, 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'affidabilità 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 verranno 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

# Elenco delle figure

1.1	An example of segmented defects from the "Severstal steel defect dataset". credits: Neven Robby and Goedemé Toon, 2021, A Multi-Branch U-Net for Steel Surface Defect Type and Severity Segmentation. <a href="https://www.mdpi.com/2075-4701/11/6/870">https://www.mdpi.com/2075-4701/11/6/870</a> . . . . .	2
1.2	Esempio di architettura a solo decoder. . . . .	4
1.3	Esempio di architettura con encoder e decoder. . . . .	5
1.4	Un diagramma di ven che illustra le relazioni tra i diversi sottogruppi dell'intelligenza artificiale, vediamo infatti come il deep learning sia un sottogruppo del representation learning, che a sua volta è un sottogruppo del machine learning. credits: Yoshua Bengio, Ian J. Goodfellow, Aaron Courville 2015, From the book "Deep Learning" . . . . .	6
1.5	Schema di un neurone biologico. . . . .	7
1.6	Esempio di rete neurale feedforward. In tale rete è possibile vedere i neuroni rappresentati dai nodi del grafo, e le interconnessioni tra di essi che definiscono il peso o l'importanza di tale connessione. File:Rete-Neurale2.svg. In Wikipedia. <a href="https://commons.wikimedia.org/wiki/File:Rete-Neurale2.svg">https://commons.wikimedia.org/wiki/File:Rete-Neurale2.svg</a> . . . . .	8
1.7	Esempio di Neurone artificiale. . . . .	8