LA SEGMENTAZIONE

INDICE:

CHE COSA È?

La segmentazione è una delle fasi che costituiscono la computer vision.
Si tratta di un processo fondamentale e problematico che consente di partizionare un'immagine in regioni disgiunte e omogenee che corrispondono ad oggetti o a parti di oggetti rappresentati in essa. Tale processo avviene attraverso l'identificazione di gruppi di pixel "omogenei", ossia pixel con caratteristiche comuni.

CHE COSA FA?

Lo scopo della segmentazione è semplificare e/o modificare la rappresentazione delle immagini per risolvere in modo efficace problemi e criticità che richiedono informazioni dettagliate sugli oggetti presenti in un'immagine, dettagli che non possono essere forniti classificando l'intera immagine o parti di essa corrispondenti ad un semplice riquadro.

Ci sono due approcci alla segmentazione:

- Per discontinuità: si ha, ad esempio, quando i bordi di una regione di interesse sono sufficientemente diversi l'uno dall'altro e anche dallo sfondo;
- Per similarità: si ha quando le regioni, anche se distanti, sono simili.

STORIA

Fino ai primi anni '80 la ricerca in analisi delle immagini si è concentrata sulle tecniche per estrarre il contorno degli oggetti, il cosiddetto edge detection, tramite tecniche di thresholding. Queste tecniche, definite come "bottom-up", vanno a segmentare immagini a partire dalle caratteristiche di basso livello, a prescindere dal contenuto semantico dell'immagine stessa.
Successivamente, si è sviluppato un interesse crescente per le tecniche "top-down", che utilizzavano modelli o conoscenze di alto livello dell'immagine per ottenere una segmentazione più precisa identificando oggetti specifici nella scena.

APPLICAZIONI

Nella realtà odierna le immagini hanno un'importanza sempre maggiore in diversi campi, di conseguenza gli use-cases della segmentazione delle immagini sono molteplici, ad esempio viene usata per:

1. La sfocatura dello sfondo (blurred background) per ottenere un effetto denominato Bokeh

2. Le auto a guida autonoma (self-driving cars) che devono poter comprendere l'ambiente che le circonda e quindi identificare linee, strade e altre informazioni essenziali.
3. La medicina e l'imaging medico, in cui la segmentazione delle immagini è utilizzata per identificare e isolare specifiche strutture anatomiche dagli scatti di imaging medico, ciò aiuta i medici nella diagnosi e nel trattamento di patologie.

TECNICHE

La segmentazione di un'immagine non è univoca, perché a seconda dell'algoritmo che si usa e delle caratteristiche che si vogliono cercare, si possono ottenere diverse segmentazioni.
Esistono molte tecniche per ottenere la segmentazione delle immagini, le quali si dividono in tradizionali e di deep learning.

THRESHOLDING

Il thresholding è un semplice metodo per segmentare un'immagine.
Esso separa gli oggetti in diverse regioni in base ad alcuni valori di soglia ed è un modo molto semplice per segmentare oggetti diversi, ad esempio, utilizzando i valori dei pixel. I valori di pixel sono diversi per gli oggetti e per lo sfondo dell'immagine se c'è un forte contrasto tra loro. In questo caso si può adottare la tecnica della soglia globale o global threshold, ossia viene scelto un valore per separare le due regioni e i pixel con valori di intensità superiori alla soglia vengono assegnati alla regione in primo piano e quelli al di sotto della soglia alla regione di sfondo. Questo metodo è semplice ed efficiente ma potrebbe non funzionare bene per immagini con illuminazione o contrasto variabili. In questi casi, le tecniche di soglia adattiva o local threshold potrebbero essere più appropriate. La soglia adattiva prevede la selezione di un valore di soglia per ciascuna regione o blocco più piccolo, in base alle statistiche dei valori dei pixel all'interno di quel blocco. La soglia adattiva è comunemente utilizzata nella scansione di documenti, nella binarizzazione delle immagini e nella segmentazione delle immagini.
Esiste una terza tecnica di thresholding, ossia la soglia di Otsu. Questa tecnica è in grado di determinare automaticamente un valore di soglia ottimale in base all'immagine che si vuole segmentare. Per questo, massimizza la varianza interclasse tra i valori dei pixel di primo piano e di sfondo dell'immagine. Quando, invece, non c'è una differenza significativa nella scala di grigi o una sovrapposizione dei valori dei pixel in scala di grigi, diventa molto difficile ottenere segmenti precisi.

Generalmente, l' immagine di input è sui toni di grigio mentre l' output è una immagine binaria in cui i pixel sono valutati come 0 o 1 . I pixel con valori al di sopra di una certa soglia vengono considerati parte del primo piano (valore 1), mentre quelli al di sotto diventano lo sfondo (valore 0).
Data un'immagine f(x,y) in bianco e nero e una soglia di intensità T. Si può calcolare l'immagine di soglia g(x,y) determinando per ogni pixel (x,y) il valore. I pixel 1 corrispondono agli oggetti dell'immagine (object points) mentre i pixel 0 allo sfondo (background points).
Il risultato dipende dalla scelta del parametro T che può essere sempre costante (soglia globale) oppure variare da pixel a pixel (soglia locale).

VANTAGGI: Non richiede una pre-elaborazione complicata.
SVANTAGGI: Presenti possibili errori di soglia.

EDGE DETECTION

L’edge detection è una tecnica di rilevamento dei bordi di un oggetto. I bordi sono discontinuità improvvise in un'immagine, che possono derivare dalla normale superficie, dal colore della superficie, dalla profondità, dall'illuminazione o da altre discontinuità.

Se l’immagine è monocromatica un edge viene definito come una discontinuità nel livello di grigio, e quindi è rilevato dove sono presenti differenze significative di luminosità. Nelle immagini a colori, invece, le informazioni per la misurazione dei bordi sono molto più compatte: per esempio, si può ricavare un edge tra due zone con stessa luminosità ma tinta differente. Nelle immagini a colori un edge viene quindi definito come una interruzione locale calcolata nello spazio colore tridimensionale.

Per rilevare i bordi si utilizzano filtri all'immagine oppure una matrice di convoluzione ma esistono anche due tecniche di segmentazione comuni, ossia:

- Rilevamento dei bordi Canny: utilizza un algoritmo a più fasi per rilevare i bordi in un'immagine. Il metodo prevede lo smussamento dell'immagine utilizzando un filtro gaussiano (eliminando un pò di rumore), il calcolo dell'ampiezza del gradiente, ossia la direzione e la velocità del cambiamento di colore, l'applicazione di una *non maxima suppression* per assottigliare i bordi e l'utilizzo della soglia di isteresi per rimuovere i bordi deboli.
- Rilevamento dei bordi Sobel: utilizza un approccio basato sul gradiente per rilevare i bordi in un'immagine. Il metodo prevede l’utilizzo dell’operatore di Sobel, un tipo di kernel di convoluzione che calcola le derivate parziali di un’immagine rispetto alle direzioni orizzontali e verticali. in seguito, si calcola l’entità del gradiente combinando i risultati ottenuti precedentemente. Infine si calcola anche la direzione del gradiente usando l'operatore di Sobel. Questo fornisce informazioni su dove avvengono i cambiamenti più significativi di intensità.

VANTAGGI: Buono per le immagini che hanno una migliore qualità contrasto tra gli oggetti.
SVANTAGGI: Non adatto per immagini rumorose.

REGION-BASED SEGMENTATION

La segmentazione basata su regioni è una tecnica utilizzata per dividere un'immagine in regioni in base a criteri di somiglianza, come colore, trama o intensità. Il metodo prevede il raggruppamento dei pixel in regioni in base alla loro somiglianza e quindi l'unione o la suddivisione delle regioni fino al raggiungimento del livello di segmentazione desiderato.

Le regioni in cui le immagini vengono suddivise devono soddisfare alcune proprietà:

- DISTINCT: nessun pixel è condiviso da due regioni
- COMPLETE: tutti i pixel dell’immagine sono assegnati ad almeno una regione della partizione
- CONNECTED: tutti i pixel appartenenti ad una regione sono “connessi”
- HOMOGENEOUS: tutte le regioni sono omogenee rispetto ad un criterio fissato (es. intensità, colore, texture, ecc..)

La segmentazione basata sulle regioni ha due tipi di approcci:

- Bottom-up. Si parte dall’ipotesi iniziale che le regioni siano costituite da un singolo pixel e, generalmente in modo iterativo, si accorpano regioni finché un certo vincolo non viene raggiunto.

- Top-down. La procedura viene inizializzata con una sola regione, contenente tutti i pixel dell'immagine; si opera quindi una suddivisione, in modo ricorsivo, fino al raggiungimento di un certo vincolo.

La segmentazione basata sulle regioni è composta da diversi metodi che utilizzano approcci differenti per identificare e raggruppare le regioni omogenee in un'immagine. Tra questi vi sono:

Region Growing: Ad ogni istante viene preso in considerazione un pixel che non è stato ancora allocato ma che è adiacente ad almeno una regione; il pixel è allocato alla regione adiacente che risulta più simile secondo il criterio scelto.

- Region Growing Algorithm:
    1. Scegli un pixel iniziale
    2. Seleziona i pixel vicini (connessi) e fai il merge se la condizione di omogeneità scelta è soddisfatta
    3. Se la regione non cresce, seleziona un altro pixel e ripeti dal punto 2 finché tutti i pixel non sono stati allocati ad una regione, altrimenti vai al punto 4
    4. Rimuovi le regioni molto piccole (passo opzionale)
- Seeded Region Growing Algorithm:
    - Viene dato in input il numero di seeds (pixel di partenza utilizzati per far crescere le regioni). L'algoritmo quindi procede autonomamente facendo crescere simultaneamente le regioni, finché tutti i pixel nell'immagine sono stati racchiusi in una regione. Per ogni passo tutti i pixel che non sono stati ancora allocati, ma che hanno almeno un vicino allocato, vengono presi in considerazione: tra tutte le regioni confinanti al pixel considerato, l'algoritmo seleziona quella i cui pixel hanno in media la minore differenza (es: in termini di livelli di grigio) rispetto al pixel preso in considerazione.

Region merging: A partire da un pixel detto "seed" si agglomerano ad esso i pixel a lui vicini che soddisfano un certo criterio di omogeneità formando così una regione. Combinando successivi processi di growing, o procedendo con growing simultaneo da più seed, si ottiene la segmentazione dell'intera immagine.

Region Splitting and Merging: È possibile eseguire una segmentazione partizionando (splitting) ricorsivamente una immagine, fino ad ottenere componenti uniformi. Si dovrà effettuare una successiva operazione di aggregazione (merging) delle regioni adiacenti che dovessero risultare compatibili in base ad un criterio di fusione.
La suddivisione ricorsiva dell'immagine in quadranti viene rappresentata con una struttura ad albero chiamato quad-tree: ogni nodo contiene le informazioni relative a ciascun quadrante e i suoi figli sono associati ai quadranti in cui è ulteriormente suddiviso. Un nodo foglia è un quadrante sufficientemente uniforme da non richiedere ulteriori partizionamenti. Dopo la fase di splitting si procederà alla fase di merging delle regioni adiacenti "compatibili"; regioni adiacenti verranno aggregate in una unica regione se quest'ultima risulterà sufficientemente uniforme.

Graph based segmentation: È una tecnica utilizzata nell'elaborazione delle immagini per dividere un'immagine in regioni in base ai bordi o ai confini tra le regioni. Il metodo prevede la rappresentazione dell'immagine come un grafico, in cui i nodi rappresentano i pixel e i bordi rappresentano la somiglianza tra i pixel. Il grafico viene quindi suddiviso in regioni minimizzando una funzione di costo, come il taglio normalizzato o l'albero di copertura minimo.

Laplacian of Gaussian edge detection: È un metodo per il rilevamento dei bordi che combina lo smussamento gaussiano con l'operatore laplaciano. Il metodo prevede l'applicazione di un filtro gaussiano all'immagine per rimuovere il rumore e quindi l'applicazione dell'operatore laplaciano per evidenziare i bordi. Il rilevamento dei bordi LoG è un metodo robusto e accurato per il rilevamento

dei bordi, ma è costoso dal punto di vista computazionale e potrebbe non funzionare bene per immagini con bordi complessi.

VANTAGGI: Funziona bene per immagini con una notevole quantità di rumore.
SVANTAGGI: Richiede tempo e memoria.

CLUSTERING

Una procedura più elaborata per effettuare l'image segmentation è quella di utilizzare la tecnica del clustering. Questa tecnica consiste in un processo di raggruppamento in sottoinsiemi, di oggetti fisici o astratti con caratteristiche analoghe.

Un cluster è una raccolta di pixel affini tra loro, che sono dissimili rispetto ai pixel degli altri cluster. Il problema degli algoritmi di clustering è decidere la misura di similarità. Non esiste una misura migliore ma ne esiste una specifica per ogni situazione.

Dalla rappresentazione dell'immagine nell'usuale spazio bidimensionale, si passa ad uno spazio delle caratteristiche, ad esempio lo spazio RGB, e si procede ad un partizionamento di tale spazio, allo scopo di definire le regioni.

La segmentazione ottenuta con il clustering non considera la relazione spaziale tra i pixel, e quindi si possono generare regioni non connesse. È necessaria un'ulteriore fase in cui tener conto delle informazioni spaziali per formare le regioni: pixel adiacenti, appartenenti alla stessa classe, costituiscono una regione omogenea. Il numero di partizioni può essere o impostato a priori o scelto in modo automatico.

L'obiettivo degli algoritmi di clustering, è quindi di determinare il raggruppamento caratteristico di un insieme di dati. Non esiste un criterio che sia in assoluto il migliore per capire se un raggruppamento sia ottimale o meno. È l'utente che applica l'algoritmo che deve stabilire quale sia il criterio che più si adatti allo scopo.

DEEP LEARNING BASED-METHODS

Image segmentation deep learning è fondamentale nel computer vision, con applicazioni diverse come le automobili a guida autonoma e l'analisi delle immagini mediche.

L'approccio del deep learning alla segmentazione delle immagini
Per la segmentazione delle immagini, il deep learning è un'ottima tecnica. Gli algoritmi di deep learning estraggono automaticamente le caratteristiche dai dati, che possono essere utilizzate per segmentare. I modelli di deep learning possono apprendere caratteristiche complesse difficili da specificare manualmente.

Le reti neurali convoluzionali (CNN), le reti completamente connesse (FCN) e le reti neurali ricorrenti (RNN) sono tra i progetti di deep learning che possono essere utilizzati per la segmentazione delle immagini. Ogni architettura presenta una propria serie di vantaggi e svantaggi.

Le reti neurali convoluzionali sono particolarmente adatte alle attività di segmentazione delle immagini, perché possono apprendere caratteristiche direttamente dalle immagini. Una CNN per la segmentazione delle immagini utilizza strati convoluzionali per estrarre caratteristiche visive , come linee, forme o texture, dalle varie regioni dell'immagine. Successivamente l' output della CNN solitamente viene elaborato ulteriormente tramite strati completamente connessi o trasversali per

associare ciascun pixel dell'immagine a una specifica classe o categoria. Questo processo consente alla rete di distinguere e assegnare etichette ai vari segmenti dell'immagine basandosi sulle caratteristiche estratte .

Per le attività di deep learning sulla segmentazione semantica delle immagini, le reti neurali ricorrenti sono un'altra opzione standard. Gli RNN sono adatti all'elaborazione di dati di serie temporali come i fotogrammi video poiché analizzano gli input in sequenza. Le dipendenze a lungo termine possono anche essere apprese dagli RNN, il che è utile per comprendere come le caratteristiche di un'immagine cambiano nel tempo.

Come funziona l'approccio del deep learning?
Una rete neurale viene utilizzata nella tecnica di segmentazione delle immagini di deep learning per imparare come dividere un'immagine in segmenti. Per addestrare la rete viene utilizzato un set di dati di immagini annotate e ciascuna immagine è etichettata con la corretta segmentazione. In questo modo, la rete apprende come mappare le foto in arrivo nelle segmentazioni appropriate.

La rete può quindi essere utilizzata per segmentare nuove immagini dopo essere stata addestrata. La rete fornirà un deep learning di segmentazione semantica per ogni nuova immagine che potrà essere utilizzata per il riconoscimento di oggetti, l'analisi di immagini mediche o qualsiasi altra applicazione.

Il tipo di rete neurale utilizzata per la segmentazione delle immagini dipende dall'applicazione. Ad esempio, una rete completamente convoluzionale (FCN) è particolarmente adatta per lavori di segmentazione delle immagini che richiedono elevata precisione.

VANTAGGI: facile implementazione grazie a librerie già pronte disponibili.
SVANTAGGI: L'addestramento del modello richiede tempo e risorse.

TIPOLOGIE

La segmentazione delle immagini può essere suddivisa nelle seguenti categorie: segmentazione delle istanze, segmentazione semantica e segmentazione panottica.

a. **Segmentazione delle istanze**

La segmentazione delle istanze prevede il rilevamento di ciascun oggetto in un'immagine, con il compito aggiuntivo di segmentare i confini dell'oggetto. L'algoritmo non ha idea della classe della regione, ma separa gli oggetti sovrapposti. La segmentazione delle istanze è utile nelle applicazioni in cui è necessario identificare e tenere traccia dei singoli oggetti.
Ad esempio, in un'immagine con quattro persone, tutte sono identificate come "persone", ma ciascuna viene trattata come un'istanza separata, considerando le differenze di altezza, colore della pelle, età e genere.

b. **Segmentazione semantica**

Durante un'attività di segmentazione semantica, le maschere di segmentazione rappresentano immagini completamente etichettate, dove tutti i pixel dell'immagine devono appartenere a una categoria, anche se non fanno parte della stessa istanza.
In questo caso, i pixel appartenenti alla stessa categoria vengono rappresentati come un unico segmento: ad esempio, tutti i pixel categorizzati come "persone" avranno lo stesso valore nella maschera di segmentazione, indipendentemente dalla loro appartenenza a diverse istanze.

c. **Segmentazione panottica**

La segmentazione panottica è una combinazione sia della segmentazione delle istanze che della segmentazione semantica. Implica l'etichettatura di ciascun pixel con un'etichetta di classe e l'identificazione di ciascuna istanza dell'oggetto nell'immagine. Questa modalità di segmentazione delle immagini fornisce la massima quantità di informazioni granulari di alta qualità provenienti da algoritmi di apprendimento automatico. La segmentazione panottica è un compito complesso di visione artificiale che risolve insieme sia i problemi di segmentazione delle istanze che quelli di segmentazione semantica. È ampiamente utilizzato, ad esempio, nei veicoli autonomi poiché le telecamere su di essi dovrebbero fornire informazioni complete sull'ambiente circostante.

SPIEGAZIONE ESPERIMENTO

GLOBAL THRESHOLDING:

1. Diamo in input l'immagine originale, il programma va a controllare se essa è in scala di grigi oppure no. Nel caso in cui l'immagine non fosse in tonalità di grigi, la si converte.
2. Successivamente, si sceglie un valore di soglia fissa. Noi abbiamo chiesto a ChatGPT che valore usare e ci ha consigliato 20, ma l'output che ci ha fornito il programma non concordava con l'esito che ci aspettavamo. Così abbiamo fatto diversi tentativi, fino ad arrivare al valore 95 che per la nostra immagine originale riteniamo essere un valore piuttosto ottimale.
3. Una volta scelto il valore della soglia abbiamo applicato il global thresholding.
4. Infine, il programma ci restituisce come output le due immagini.

SOBEL EDGE:

1. L'immagine originale viene presa in input dal programma e viene convertita in scala di grigi.
2. All'immagine convertita viene applicato il filtro di Sobel per identificare i bordi presenti sull'asse delle *x* e delle *y*.
3. Viene poi applicata la convoluzione con i filtri di Sobel. Una volta applicata la convoluzione, viene calcolato, in un primo momento, il gradiente totale, per poi essere normalizzato in modo che potesse assumere un valore compreso tra 0 e 255.
4. Il programma ci restituisce come output l'immagine di partenza a paragone con l'immagine calcolata attraverso il Sobel Edge Detection.

REGION GROWING:

1. Diamo in input al programma l'immagine, il programma va a controllare se l'immagine è a scala di grigi, se non lo è verrà convertita.
2. Successivamente si inizializza la maschera e la coda di pixel da esaminare.
3. A questo punto inizia il vero e proprio region growing: si va a creare un array di pixel in cui si andranno a mettere tutti i pixel adiacenti. Per capire se un pixel va messo in questo array, bisogna vedere se è già stato visitato, se non è stato ancora visitato bisogna calcolare la differenza del valore di questo pixel con il valore del pixel a lui adiacente e se questa differenza è inferiore alla soglia allora il pixel viene aggiunto all'array di pixel adiacenti. In questo modo si crea la regione.
4. Il risultato ci viene fornito in output dal programma.

NOI COME SEGMENTIAMO?

Noi sappiamo che la segmentazione è un problema difficile, quindi la domanda è: "come possiamo realizzarlo noi essere umani?".

Per comprendere come gli esseri umani realizzano la segmentazione, bisogna collegarsi alla teoria della Gestalt, sviluppata all'inizio del Novecento da un gruppo di scienziati tedeschi. Essa afferma che ciò che facciamo è percepire gli oggetti nella loro interezza e, una volta fatto ciò, siamo in grado di raggruppare l'intero oggetto, solo dopo identifichiamo le sue singole parti o sottogruppi.

La segmentazione delle immagini è altamente soggettiva per gli essere umani. Ciò lo si nota da un esperimento condotto da David Martin all'università di Berkeley nel 2001, in cui venne chiesto a tre persone diverse di trovare regioni o segmenti significativi in un'immagine rappresentante una chiesa. Si notò che le prime due persone avevano segmentato in modo abbastanza simile, invece la terza aveva aggiunto maggiori dettagli distinguendosi dagli altri utenti.

"*It is established since the Gestalt movement in psychology that perceptual grouping plays a fundamental role in human perception*." (R. Nock – F. Nielsen)

La percezione visiva umana è governata dai principi della Gestalt e molte strategie di segmentazione sono riconducibili ad essi.
Possiamo notare che ci sono delle caratteristiche comuni tra i vari algoritmi di segmentazione e i principi della Gestalt, ad esempio c'è una correlazione tra:

- Il principio di figura/sfondo e il thresholding, in quanto entrambi vanno a distinguere tra due regioni, lo sfondo e gli oggetti.
- Il principio di similarità con il clustering, in quanto entrambi raggruppano oggetti con caratteristiche analoghe
- Il principio di prossimità con il region growing e il region merging, in quanto entrambi prendono oggetti o pixel adiacenti
- I principi di continuità e chiusura con l'edge detection in quanto entrambi tendono a proseguire lungo una linea e a completare linee non chiuse

CONCLUSIONI

Al termine della nostra analisi, emerge chiaramente che il processo umano di segmentazione delle immagini si basa su intuizioni e esperienze, complicando notevolmente la traduzione di questo complesso processo cognitivo in un singolo algoritmo. La sfida risiede nell'incorporare la ricchezza delle nostre intuizioni e nella comprensione contestuale in un modello computazionale unico e preciso.