

Riconoscimento delle Targhe con il Machine Learning su AWS: un'analisi approfondita di una soluzione reale

1 Settembre 2023 - 11 min. read

[Amazon Rekognition](#)

[Amazon Textract](#)

Overview

Nel panorama tecnologico in rapida evoluzione di oggi, la convergenza tra intelligenza artificiale e cloud computing ha aperto nuove vie per risolvere sfide reali. Tra queste rientra sicuramente il dominio del Riconoscimento Ottico dei Caratteri (OCR). Superando il semplice riconoscimento del testo, l'evoluzione dell'OCR ha abilitato la capacità di estrarre e decifrare con precisione informazioni diventando un pilastro per l'automazione e l'efficienza operativa.

Una delle applicazioni più interessanti dell'OCR degli ultimi anni è stata in relazione all'**elaborazione della targhe automobilistiche**. Per citare qualche esempio, si pensi alle imprese il cui core business ha a che fare con un grande numero di veicoli: in un parcheggio con alto flusso di veicoli in entrata e in uscita, l'integrazione di questa tecnologia abilita l'elaborazione automatica dei pagamenti ed elimina la necessità di transazioni manuali.

Analogamente, per una stazione di servizio sarebbe possibile realizzare un sistema di pagamento automatico del rifornimento e, contemporaneamente, raccogliere statistiche sul traffico e sui potenziali clienti.

Altro caso interessante è l'applicazione nel settore pubblico in cui, per esempio, le agenzie di forze dell'ordine possono sfruttare il potere dell'OCR per identificare rapidamente i veicoli di interesse. Altre applicazioni possono spaziare dalla gestione del traffico, alla raccolta di informazioni sull'affluenza ad eventi, fino al monitoraggio del traffico veicolare presso zone designate. Inoltre, gli urbanisti possono utilizzare questa tecnologia per ottenere preziose informazioni sui modelli di traffico, rendendo la pianificazione urbana più basata sui dati e reattiva. Anche le autorità responsabili della manutenzione delle strade

potrebbero gestire in modo più efficiente la riscossione dei pedaggi e monitorare l'uso delle autostrade, migliorando la generazione di ricavi e la gestione delle infrastrutture stradali.

Si tratta solo di alcuni esempi, ma le applicazioni sono davvero variegata e possono spaziare fino ad arrivare in ambiti in cui il veicolo non è nemmeno il punto focale del servizio, ma diventa solo un modo per riconoscere un utente e migliorare la sua esperienza nei confronti di un brand o un servizio.

Questo articolo vuole essere un diario della nostre ricerche in cui abbiamo affrontato tutte le complessità dell'OCR applicato alle targhe, ovviamente trovando la miglior soluzione in ambiente AWS. Il nostro viaggio ci ha portato a considerare pro e contro di alcuni servizi fino a determinare la soluzione più efficace per l'estrazione accurata del testo delle targhe automobilistiche all'interno di un caso d'uso reale. Il nostro obiettivo, dunque, è fornire non solo la conoscenza dei servizi AWS per l'OCR al fine di permettervi di selezionare il migliore per il vostro specifico caso, ma dare una comprensione completa dei potenziali ostacoli e di come affrontarli.

Siete pronti? Iniziamo!

OCR on AWS

All'interno dell'ampia gamma di servizi offerti da Amazon Web Services (AWS), è disponibile una varietà di servizi di Intelligenza Artificiale (AI) e Machine Learning (ML) utili a costruire il nostro sistema di decodifica delle targhe. Tra questi, Amazon Textract, Amazon Rekognition e SageMaker giocano un ruolo chiave.

Per il nostro specifico progetto, però, la direttiva era chiara: dare priorità all'utilizzo dei servizi gestiti all'interno dell'ecosistema AWS. Di conseguenza, la nostra attenzione si è concentrata esclusivamente su Amazon Textract e Amazon Rekognition.

Introduciamo formalmente entrambi questi servizi prima di proseguire nel nostro viaggio.

Amazon Textract

Amazon Textract, un componente eccezionale del repertorio di AWS, è un sofisticato servizio di ML progettato per estrarre testo e dati da diversi documenti, immagini e moduli. La sua forza principale risiede nella sua capacità di elaborare in modo accurato ed efficiente grandi volumi di informazioni testuali, trasformando contenuti non strutturati in dati strutturati con notevole precisione. Questo servizio dimostra un'eccezionale adattabilità, eccellendo in vari scenari come l'elaborazione delle fatture, moduli e tabulati di vario tipo.

Le caratteristiche più rilevanti includono la sua capacità di distinguere tra diversi tipi di contenuti all'interno dei documenti, come tabelle, moduli e blocchi di testo. Questa granularità di analisi migliora la sua capacità di estrarre informazioni pertinenti in modo efficace. La sua integrazione con i servizi AWS consente il trasferimento dei dati in modo uniforme per ulteriori analisi e applicazioni. Inoltre, l'identificazione di coppie chiave-valore e del loro contesto è preziosa, offrendo una visione completa dei dati strutturati.

Amazon Textract emerge subito come una soluzione convincente in grado di ottimizzare significativamente gli obiettivi del nostro progetto.

Amazon Rekognition

In prima linea tra le offerte di AI e ML di AWS, Amazon Rekognition è un formidabile servizio di analisi di immagini e video progettato per estrarre informazioni da contenuti visivi. La sua potenza risiede nella sua capacità non solo di identificare oggetti, volti e scenari all'interno di immagini e video, ma anche di rilevare e riconoscere il testo presente in questi elementi visivi. Questo lo rende un asset di grande valore nel dominio del Riconoscimento Ottico dei Caratteri (OCR).

Le caratteristiche più significative di Amazon Rekognition includono le sue avanzate capacità di analisi dei volti, che consentono la rilevazione di diversi aspetti tra cui le emozioni e persino del riconoscimento completo di un viso mediante confronto o cercando in un database precaricato per determinarne l'identità. Inoltre, la sua abilità nel rilevare il testo va oltre la semplice identificazione, permettendo di catturare dettagli di alta precisione all'interno delle immagini. L'integrazione della funzione di labeling personalizzabile aumenta ulteriormente la sua applicabilità in scenari come la moderazione dei contenuti e la conformità alla sicurezza.

La sua versatilità nel gestire casi d'uso diversificati e la sua integrazione seamless con i servizi AWS ne sostengono il potenziale per contribuire in modo significativo agli obiettivi del nostro progetto.

Con l'introduzione di questi servizi completata, possiamo ora procedere senza indugi con il nostro viaggio.

Step 1: Preprocessing dell'immagine

Indipendentemente dal servizio scelto, diventa essenziale adottare alcuni step di preprocessing. Nel nostro caso, abbiamo optato per sfruttare le capacità di Amazon Rekognition per agevolare diverse fasi preliminari, in particolare l'estrazione del bounding box che racchiude la targa. Tuttavia, il percorso che ci attendeva era tutt'altro che diretto. Le foto provenienti dal campo in situazioni reali sono spesso catturate da angolazioni diverse, in condizioni di illuminazione varie e contengono una vasta gamma di veicoli

diversi, con formati di targa profondamente disomogenei. Ciò ha comportato sfide tutt'altro che banali. Il tentativo di acquisire con precisione il bounding box della targa si è rivelato un'impresa più ardua del previsto.

La semplice presentazione dell'immagine originale ad alta risoluzione si è rivelata insufficiente non permettendoci di identificare la targa in un numero di casi considerato accettabile. Per arrivare ad un risultato soddisfacente, oltre all'utilizzo di Rekognition abbiamo dovuto impiegare alcune tecniche di preelaborazione delle immagini per migliorare l'accuratezza dell'output.

Ci siamo concentrati, dunque, sulla manipolazione dei colori. Riconoscendo l'importanza della loro rappresentazione ottimale, abbiamo strategicamente convertito le immagini in **scala di grigi**. Questa scelta strategica ha semplificato le analisi successive e ridotto le possibili interferenze dalle variazioni cromatiche dovute a diverse condizioni di illuminazione.

Successivamente, la nostra attenzione si è spostata sull'accentuazione del contrasto, una tecnica fondamentale nell'aumentare la chiarezza degli elementi visivi. Ci siamo quindi addentrati nel campo della **posterizzazione**, un metodo che ha suddiviso l'immagine in scala di grigi in regioni distinte e visivamente discernibili. In tal modo, abbiamo mirato ad accentuare il confine tra la targa e il suo contesto, agevolando il suo isolamento e riconoscimento.

Portando la ricerca della chiarezza ancora oltre, abbiamo introdotto un processo di **dithering**.

Questa tecnica prevedeva l'organizzazione dei pixel per simulare sfumature e sfaccettature aggiuntive, colmando efficacemente il divario tra scala di grigi e una rappresentazione più ricca e dettagliata.

Attraverso questa fusione di tecniche siamo riusciti a ottenere un'immagine raffinata che mantenesse l'essenza dell'originale, minimizzando al contempo l'impatto del rumore visivo.

Abbiamo quindi proceduto a sottoporre un notevole volume di immagini così processate a Rekognition, scoprendo un fenomeno interessante: alcune immagini in cui la targa veniva riconosciuta con successo nel formato originale non producevano risultati accurati una volta elaborate, mentre alcune immagini in cui la targa non era stata rilevata inizialmente, venivano ora riconosciute correttamente.

Tale punto ha richiesto un approfondimento.

Alla luce di un'analisi più accurata, è emersa un'idea: **la riduzione del rumore, sebbene apparentemente vantaggiosa, non migliorava in modo coerente la capacità di Rekognition**

di identificare le targhe. Ciò accadeva perché il modello sottostante di Rekognition non è esclusivamente tarato per localizzare le targhe. Invece, il modello utilizza criteri diversi per determinare se una targa è presente in un frame - criteri che la nostra preelaborazione ha involontariamente modificato, evidenziando solo le lettere ad alto contrasto.

Alla luce di ciò, abbiamo ricalibrato la nostra strategia come segue:

Sottoposizione a Rekognition dell'Immagine originale: Il primo step è presentare l'immagine non modificata ad alta risoluzione a Rekognition per ottenere il bounding box della targa.

Nei casi in cui l'analisi iniziale di Rekognition non riusciva a rilevare una targa, avremmo applicato le fasi di pre elaborazione dettagliate in precedenza. La versione migliorata dell'immagine veniva quindi nuovamente sottoposta a Rekognition per un nuovo tentativo di rilevamento. A questo punto la stragrande maggioranza delle immagini veniva correttamente riconosciuta.

Una volta che Rekognition identificava con successo il bounding box della targa, procedevamo alla fase successiva.

Step 2: Estrazione del testo dalle targhe

Procediamo a calcolare un bounding box leggermente più ampio di quello inizialmente identificato da Rekognition. Questa espansione strategica assicura un'incapsulazione completa della targa, garantendo così ampio spazio contestuale per consentire ad altri servizi basati sull'IA di verificare la presenza di una targa. Questo allargamento intenzionale **protegge da possibili perdite di informazioni** mantenendo un quadro visivo equilibrato.

Procediamo quindi al ritaglio dell'immagine e all'applicazione delle procedure di pre elaborazione descritte in precedenza, a meno che l'immagine non abbia subito un trattamento precedente.

Il risultato del ritaglio dell'immagine presenta una composizione convincente: una targa al centro, caratterizzata da un rumore visivo minimo e da un contrasto accentuato, ed è pronta per essere inviata al servizio OCR scelto.

Di primo acchito, la nostra attenzione si è concentrata sullo sfruttamento delle capacità di Amazon Textract, che abbiamo già avuto modo di testare, ad esempio in [questo articolo](#). Le nostre osservazioni sulle prestazioni di Amazon Textract nel trattare vari documenti hanno sottolineato le sue capacità robuste. Ciò ci ha portato a una deduzione logica: con l'implementazione di passaggi di pre elaborazione mirati, il servizio poteva essere adattato al compito specifico di estrarre testo dalle targhe automobilistiche.

La nostra ipotesi si basava sul presupposto che presentare a Textract immagini incontaminate e ad alto contrasto avrebbe prodotto risultati simili alle sue prestazioni con i documenti. Così, abbiamo inviato le immagini preelaborate a Textract per l'estrazione del testo, ma anche in questo caso la realtà si è discostata dalle nostre aspettative evidenziando ulteriori complessità.

Indipendentemente da quanto venissero elaborate le immagini, le prestazioni di Textract non sono state all'altezza di produrre risultati utilizzabili. Textract è evidentemente ottimizzato per compiti orientati per lo più ai documenti.

Conseguentemente, siamo giunti a un punto cruciale, che ci ha portato a riconsiderare Rekognition come unico strumento versatile per l'estrazione del testo, una strada che ha dimostrato un potenziale significativo nonostante la nostra iniziale focalizzazione su Textract.

Abbiamo rivisitato con Rekognition le stesse immagini preelaborate precedentemente inviate a Textract ottenendo risultati promettenti.

Abbiamo testato in questo modo la soluzione utilizzando immagini con diversi gradi di complessità; mentre le immagini con condizioni ottimali venivano rilevate con precisione senza la necessità di pre-elaborazione, le immagini più intricate avevano successo solo dopo essere state sottoposte a elaborazione preliminare.

Come risultato siamo giunti alla conclusione decisiva: l'impiego dell'immagine ritagliata ad alto contrasto costituiva l'opzione più affidabile ed efficace.

La pre-elaborazione dell'immagine fortemente personalizzata ne migliora l'idoneità per l'analisi attraverso Rekognition, almeno per quanto riguarda l'estrazione del testo, e garantisce prestazioni uniformi su una gamma di complessità delle immagini.

Step 3: Post-processing dell'output

Una volta che il testo era stato estratto con successo, molti casi richiedevano un'ulteriore attenta elaborazione. Ecco alcuni esempi:

Rekognition si è rivelato incapace di differenziare in modo preciso alcune lettere tra gli alfabeti cirillico e latino. Nel testo estratto, spesso le lettere visivamente identiche venivano codificate utilizzando sequenze di bit UTF8 errate. Questo ha portato all'interpretazione di queste lettere come caratteri cirillici, generando di conseguenza un testo della targa in esame che differiva dall'immagine originale.

Purtroppo, Rekognition non dispone di un meccanismo per suggerire o forzare la scelta dell'alfabeto durante l'estrazione. Di conseguenza, è necessario un passaggio di post-

processing per rettificare questa discrepanza. La prima procedura di post-processing comporta la rimozione di tutte le lettere cirilliche, e la loro sostituzione con l'equivalente latino (dal punto di vista grafico); un'operazione resa possibile dal fatto che questi caratteri non sono mai presenti nelle targhe esaminate nel nostro caso d'uso.

Un'altra importante attività di post-processing riguarda il riassemblaggio delle targhe che si estendono su più righe, come ad esempio quelle delle motociclette. Anche in questo caso, le capacità di Rekognition si dimostrano utili. Utilizzando l'API di etichettatura, Rekognition riesce a comunicarci se un'immagine cattura un'auto o una motocicletta. Nei casi in cui viene rilevata una motocicletta, il testo estratto deve essere ricostruito e formattato su una singola riga.

Queste procedure di post-processing sottolineano l'interazione intricata tra l'estrazione guidata dall'IA e le sfumature intrinseche ai dati testuali. La **fusione tra analisi automatizzata e successiva raffinazione guidata dall'essere umano** garantisce che il testo estratto si allinei con precisione a scenari del mondo reale, producendo una soluzione affidabile e robusta per l'OCR delle targhe.

Conclusioni

In questo articolo, abbiamo presentato il nostro approccio empirico alla costruzione di una soluzione OCR per targhe sulla piattaforma AWS, sfruttando le capacità dei servizi gestiti. Abbiamo confrontato Amazon Textract e Amazon Rekognition e abbiamo valutato le loro prestazioni su un set di dati di immagini di targhe automobilistiche anche disomogenee. Abbiamo anche discusso i passaggi di preprocessing necessari, come il ridimensionamento, il ritaglio e l'ottimizzazione delle immagini, che possono migliorare l'accuratezza del processo OCR. Abbiamo constatato che in relazione a questo preciso momento storico, Amazon Rekognition si attesta come il miglior servizio per l'estrazione del testo quando si parla di targhe automobilistiche in quanto offre l'accuratezza più elevata, la latenza più bassa e la maggiore flessibilità tra i servizi AWS.

Vi siete mai trovati a usufruire di un servizio automatizzato basato sulla targa della vostra auto? Ora sapete da quali componenti (e da quali sfide vinte) si compone!

Continuate a seguirci per nuovi articoli su AI e Machine Learning.

A presto su Proud2beCloud!

About Proud2beCloud

Proud2beCloud è il blog di [beSharp](#), APN Premier Consulting Partner italiano esperto nella progettazione, implementazione e gestione di infrastrutture Cloud complesse e servizi AWS avanzati. Prima di essere scrittori, siamo Solutions Architect che, dal 2007, lavorano

quotidianamente con i servizi AWS. Siamo innovatori alla costante ricerca della soluzione più all'avanguardia per noi e per i nostri clienti. Su Proud2beCloud condividiamo regolarmente i nostri migliori spunti con chi come noi, per lavoro o per passione, lavora con il Cloud di AWS. Partecipa alla discussione!



Alessio Gandini

Cloud-native Development Line Manager @ beSharp, DevOps Engineer e AWS expert. Computer geek da quando avevo 6 anni, appassionato di informatica ed elettronica a tutto tondo. Ultimamente sto esplorando l'esperienza utente vocale e il mondo dell'IoT. Appassionato di cinema e grande consumatore di serie TV, videoggiocatore della domenica.

Copyright © 2011-2023 by beSharp spa - P.IVA IT02415160189