HS.Vision: L'analisi visuale di Humatics s.r.l.

Marco Cristani[⋆ +], **Pietro Lovato**[+], **Davide Conigliaro**[+]
[⋆]Universita degli Studi di Verona, [+]Humatics s.r.l.
marco.cristani@univr.it,pietro.lovato@sys-datgroup.com, davide.conigliaro@sys-datgroup.com

Abstract

Questo contributo presenta HS.Vision, il prodotto di punta di Humatics s.r.l., azienda del gruppo SYS-DAT Group s.p.a.. HS.Vision è un motore di inferenza per dati visuali capace di svolgere molteplici task, tutti incentrati sull'analisi di immagini in cui siano presenti persone. In particolare, verranno mostrate qui le due applicazioni principali, ovvero l'analisi automatica di filmati social e il conteggio e l'analisi di assembramenti in spazi vasti e strutturati, analizzando come casi di studio l'analisi dei video social di TikTok e Instagram e il resoconto di un anno di monitoring di 11 eventi (quali Marmomac 2021, Fieracavalli 2021) presso i quartieri di VeronaFiere s.p.a..

1 Introduzione

I contenuti visuali hanno assunto un'importanza cruciale e strategica nel mondo digitale. Siano foto caricate in rete, o *frame* acquisiti da telecamere di sorveglianza, in ogni caso le immagini richiedono una capacità di analisi nel breve periodo sempre più proibitiva.

Questo articolo presenta HS.Vision, il motore di inferenza visuale che rappresenta il prodotto di punta di Humatics s.r.l., azienda del gruppo SYS-DAT Group s.p.a.. HS.Vision è un insieme di algoritmi di analisi per le immagini progettati nel corso degli anni presso il Dipartimento di Informatica dell'Università di Verona, derivanti da pubblicazioni scientifiche [Godi *et al.*, 2022] ed ingegnerizzati presso Humatics s.r.l., dapprima spin-off accreditata dall'Universita di Verona, poi azienda acquisita da SYS-DAT Group s.p.a. nell'agosto 2021.

In particolare, questo articolo si sofferma su due applicazioni principali di HS.Vision, ovvero per l'analisi dei video social, per capire che vestiti siano indossati dalle persone riprese, e per l'analisi di filmati di sorveglianza, per contare le persone, analizzare la presenza di assembramenti ed come sviluppo corrente, il corretto uso di DPI. Nel primo caso si fa riferimento all'analisi di video di influencer su TikTok e Instagram, mostrando come, grazie ad HS.Vision, siano necessari pochi fotogrammi (10) di un video social per individuare il prodotto corretto indossato nel video da una galleria di elementi di oltre 14.000 con una precisione fino all'80%. Nel secondo caso si mostra come HS.vision sia stato capace di gestire sistemi di centinaia di telecamere funzionanti contemporaneamente durante più di un anno di continua attività per VeronaFiere, l'azienda che gestisce l'organizzazione e ospita le fiere più importanti del panorama italiano quali Vinitaly, Fieracavalli, Marmomac.

Il proseguio dell'articolo è organizzato come segue: In Sez. 2 verrà presentata brevemente Humatics s.r.l., dettagliando il suo percorso dalla nascita all'acquisizione, avvenuta solo dopo 5 anni. Successivamente in Sez. 3 verrà presentato HS.Vision nell'applicazione sui video social; in Sez. 4 verrà illustrato HS.Vision per il monitoring di vasti spazi indoor; infine, in Sez. 5 verranno presentate brevi conclusioni sugli sviluppi in corso di HS.Vision.

2 Humatics s.r.l.

Humatics nasce come start-up innovativa il 29 Febbraio 2016. I soci fondatori sono dr. Davide Conigliaro, dr. Pietro Lovato e prof. Marco Cristani, conosciutisi nel 2009, dapprima come studenti e docenti (prof. Cristani) all'interno della laurea in Ingegneria e Scienze Informatiche dell'Università di Verona, poi attraverso collaborazioni di ricerca; in questo periodo si crea un legame di fiducia ed entusiasmo che li porta ad iniziare la loro prima avventura imprenditoriale. Humatics s.r.l nasce come start-up innovativa, per passare dopo pochi mesi a spin-off accreditata dell'Universita di Verona, e vincere sempre nel 2016 la Start Cup Veneto. Grazie al ruolo di spin-off, Humatics cresce sia a livello tecnologico che di conoscenza del mercato, grazie alle opportunità offerte dal Computer Science Park (CSP) all'interno dell'Università di Verona (https://www.csp.univr.it/): il CSP consente alle aziende che ne fanno parte di avere accesso a conoscenze, tecnologie, metodi di produzione, prototipi e servizi. Questo si è concretizzato, tra l'altro, attraverso contratti di ricerca instaurati con l'Università di Verona e grazie ad attività di stage e tirocinio, dove numerosi studenti hanno contribuito al lavoro di Humatics, e si sono fatte conoscere come potenziali leve. Dopo 5 anni dalla nascita, Humatics è stata oggetto di due quotazioni interessanti, cedendo alla fine la maggioranza a SYS-DAT Group s.p.a.. Ad oggi il personale di Humatics consta di 5 membri, più 5 collaboratori.

3 HS.Vision per l'analisi dei video social

Individuare l'outfit di una celebrità o di un social influencer può trasformare i video in spot pubblicitari inestimabili, in un mercato in cui vengono caricate e visualizzate oltre un miliardo di ore di video ogni giorno [Duffett, 2020], per circa 3 ore al giorno [Koponen, 2020; Oberlo, 2020a; Oberlo, 2020b]. Si stima che il numero di utenti globali che trasmetteranno regolarmente video in streaming raggiungerà i 4,5 miliardi nei prossimi cinque anni [Smith, 2018], mostrando il potenziale dei video rispetto alle immagini generiche come strumento di marketing *general purpose* [Dopson, 2020]. Il compito di associare le immagini provenienti dai social network alle immagini contenute nei repository degli e-commerce si chiama *street-to-shop* problem [Hadi Kiapour *et al.*, 2015]. L'estensione al video viene denominata *video-to-shop*.

HS.Vision contiene una tecnica video-to-shop derivata da [Godi *et al.*, 2022], che a partire da una sequenza video "social", ne individua i prodotti e ne estrae le caratteristiche visuali, adottando un meccanismo proprietario di raccolta e aggregazione dei dettagli visuali, abbinando tali prodotti a quelli presenti in una galleria di immagini "pulite", ovvero per esempio le immagini presenti in un sito di e-commerce, quindi senza rumore su sfondo neutro. HS.Vision estende il popolare Match-RCNN [Ge *et al.*, 2019], lo stato dell'arte del problema a singola immagine street-to-shop, applicando una tecnica di domain-adaptation che permette di trasferire la conoscenza acquisita per risolvere il problema da singola immagine ai video. In pratica, l'idea è quella di capire in una sequenza quali siano i particolari associabili con certezza ad un vestito (distinguendo così dallo sfondo) che siano maggiormente discriminativi rispetto a molti prodotti. Questa idea è nota in letteratura come *analisi di attenzione*

Gli esperimenti in [Godi *et al.*, 2022] sono stati applicati a Moving Fashion, un dataset da 5.854 milioni di fotogrammi annotati, organizzati in 15045 coppie <*video social, prodotto e-commerce*>, ovvero ogni video è associato a una distinta immagine di un prodotto presente in un sito di e-commerce. I risultati mostrano come siano necessari pochi fotogrammi (10)di un video social per individuare il prodotto corretto indos-sato nel video da una galleria di elementi di oltre 14.000 con una precisione fino all'80%.

Negli esperimenti, è possibile notare che la tecnica dell'attenzione gioca un ruolo cruciale per la performance HS.Vision. A seguire illustriamo il suo ruolo qualitativamente e quantitativamente.

In Fig. 1 riportiamo i valori di attenzione catturati su ogni fotogramma della sequenza video, in espressi come numeri reali da 0 a 1. Nella riga a), si può notare che l'attenzione è alta quando il il logo del cuore è visibile (0.31, 0.23 nei primi due fotogrammi) e si abbassa quando scompare, nonostante la maglia azzurra (ultimo fotogramma) sia molto simile per area. Ciò significa che il meccanismo considera il logo del cuore importante per l'individuazione di quel prodotto. Nella seconda riga b), si vede l'effetto di un'occlusione nel punteggio di attenzione (ultimo fotogramma). Nella terza riga c), un top bianco con un logo dà un punteggio di attenzione stabile (intorno a 0,28). Se si copre manualmente il logo come

Figura 1: Osservazioni qualitative sul comportamento di attenzione. Sulla sinistra, per ogni sequenza video mostriamo i riquadri di rilevamento e il punteggio di attenzione calcolato. A destra l'articolo del negozio abbinato.

abbiamo fatto nel terzo frame, si provoca una netta diminuzione dell'attenzione, aumentando uniformemente quelle che evidenziano il logo.

4 HS.Vision per l'analisi real-time di grandi spazi

HS.Vision per il monitoring di grandi spazi è un sistema di analisi video che consente di collegarsi a qualsiasi telecamera per rilevare la presenza delle persone ed estrarre informazioni utili alla sicurezza, il tutto nel rispetto della privacy grazie ad un'analisi proprietaria secondo i principi di *privacy by design*. La sua compatibilità con qualsiasi sistema di sorveglianza già costituito abbatte i costi del prodotto e quelli di installazione.

HS.Vision per il monitoring è stato richiesto nel 2020 da Veronafiere s.p.a. come misura per la gestione dell'emergenza sanitaria COVID-19, ed in generale come sistema per connettere e gestire in maniera centralizzata le porzioni dei diversi quartieri fieristici. Lo spazio espositivo di Veronafiere consta di 309.000 mq con 13 padiglioni, in grado di ospitare decine di migliaia di visitatori contemporaneamente. Basti pensare che nel 2019 sono state organizzate 71 manifestazioni con oltre 1 Milione di visitatori. Dislocate nei vari padiglioni vi sono oltre 400 telecamere di video sorveglianza, provenienti da diversi fornitori, del tutto eterogenee. Per esempio, in Fig. 2, è riportato un esempio di scenario considerato da HS.Vision, ripreso da telecamera *fisheye*.

HS.Vision ha dato per la prima volta un controllo omogeneo di tutta questa sensoristica, grazie ad un'interfaccia web per l'analisi in tempo reale (Fig. 3), ed un sistema in grado di elaborare molteplici statistiche (flussi temporizzati in e out da un padiglione, distribuzione delle persone totali nei vari padiglioni, etc). All'interno di ogni singolo padiglione, vi è la

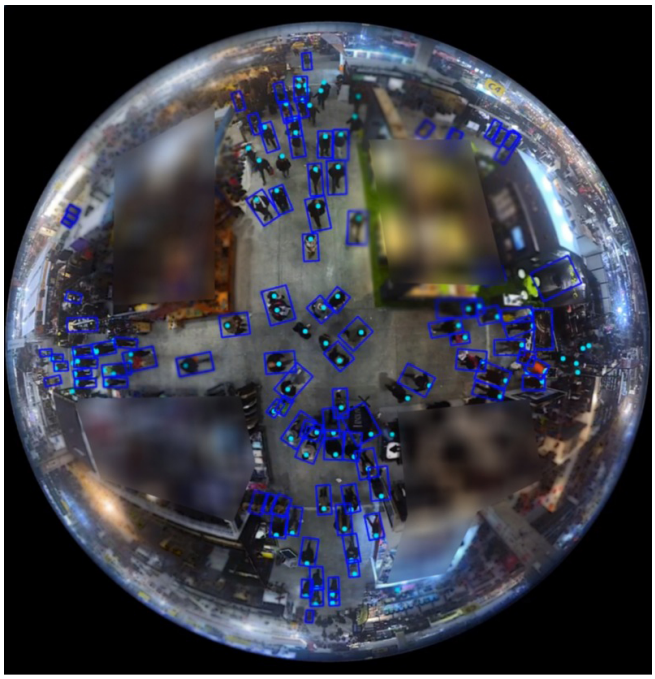

Figura 2: una telecamera fisheye operante nei quartieri di Veronafiere, con sovrimpressa una grafica che mostra il sistema di conteggio delle persone eseguito da HS.Vision.

possibilità di istanziare un'analisi a grana fine di tutti gli assembramenti occorrenti, e di allertare in tempo reale gli operatori di sicurezza, nel caso di situazioni critiche, come per esempio la presenza di un forte assembramento. Per quanto riguarda l'analisi di DPI, HS.Vision è attualmente installata in ambiente di laboratorio, dove sta applicando continuativamente la rilevazione automatica di mascherine. La sperimentazione sta dando risultati ottimi, e si prevede che la nuova funzionalità diventi parte del prodotto entro l'anno.

5 Conclusioni

Il presente articolo mostra come Humatics sia una delle realtà industriali italiane in grado di proporre prodotti realmente basati su ricerca allo stato dell'arte nell'intelligenza artificiale. Gli sviluppi di HS.Vision futuri vedranno in particolare la collaborazione con una multinazionale nell'ambito della video analisi per quanto riguarda l'applicazione della moda. Relativamente all'analisi real time di grande spazi, HS.Vision è stata adottata anche per il prossimo anno solare da parte di Veronafiere, a testimonianza dei successi ottenuti lo scorso anno, e della fiducia riposta nel prodotto e nei suoi sviluppi futuri.

Da una parte, capire come vengono percepite le proprie immagini è una sfida cruciale per le aziende che vogliono ottenere un vantaggio competitivo soprattutto nel panorama italiano, in cui livello tecnologico delle principali soluzioni di *video analytics* è spesso un passo indietro rispetto ai competitor esteri.

Figura 3: Esempio di densità di assembramenti all'interno di un quartiere fieristico, come riportato in real-time dal sistema di monitoring di HS.Vision.

Riferimenti bibliografici

[Dopson, 2020] Elise Dopson. Videos vs. images: Which drives more engagement in facebook ads?, 2020. Available online: https://databox.com/videos-vs-images-in-facebook-ads [Accessed: 10 November 2020].

[Duffett, 2020] Rodney Duffett. The youtube marketing communication effect on cognitive, affective and behavioural attitudes among generation z consumers. *Sustainability*, 12(12):5075, 2020.

[Ge *et al.*, 2019] Yuying Ge, Ruimao Zhang, Xiaogang Wang, Xiaoou Tang, e Ping Luo. Deepfashion2: A versatile benchmark for detection, pose estimation, segmentation and re-identification of clothing images. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5337–5345, 2019.

[Godi *et al.*, 2022] Marco Godi, Christian Joppi, Geri Skenderi, e Marco Cristani. Movingfashion: a benchmark for the video-to-shop challenge. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1678–1686, 2022.

[Hadi Kiapour *et al.*, 2015] M Hadi Kiapour, Xufeng Han, Svetlana Lazebnik, Alexander C Berg, e Tamara L Berg. Where to buy it: Matching street clothing photos in online shops. In *Proceedings of the IEEE international conference on computer vision*, pages 3343–3351, 2015.

[Koponen, 2020] Meri Ester Koponen. How to create engaging mobile-optimised video ads for social media., 2020.

[Oberlo, 2020a] Oberlo. 10 social media statistics that you need to know in 2020, 2020. Available online: https://www.oberlo.com/blog/social-mediamarketing-statistics [Accessed: 24 March 2020].

[Oberlo, 2020b] Oberlo. 10 tiktok statistics that you need to know in 2020, 2020. Available online: https://www.oberlo.com/blog/social-mediamarketing-statistics [Accessed: 24 March 2020].

[Smith, 2018] Kit Smith. 57 fascinating and incredible youtube statistics, 2018. Available online: https://www.brandwatch.com/ blog/youtube-stats/ (accessed on 18 March 2020).