POLITECNICO DI TORINO

Corso di Laurea
in Ingegneria Matematica

Tesi di Laurea

Modelli predittivi della saliency nelle rappresentazioni grafiche dei dati

Relatori
prof. Fulvio Corno
prof. Luigi De Russis
prof. Luisa Fernanda Barrera Leon
firma dei relatori

. . . . . .
. . . . . .
. . . . . .

Candidato
Samuel Grassi

firma del candidato

. . . . . .

Anno Accademico 2020-2021

Obiettivo

L'obiettivo di questo lavoro di tesi è quello di approfondire ed entrare nel dettaglio dell'argomento della *saliency* nell'analisi e nella distribuzione dei punti di fissazione di un osservatore su grafici di business.

Inizialmente il focus era quello di cercare di trovare un algoritmo competitivo con alcuni modelli sulla *saliency* visiva già proposti (Itti, Matzen), ma in seguito, dopo aver analizzato i risultati su un dataset già esistente (*MASSVIS*), ci si è concentrati sul valutare attentamente le performance di questi modelli già esistenti su un nuovo esperimento creato e controllato da noi.

Infine, alla luce di quanto trovato e del materiale prodotto, il lavoro si conclude con alcuni spunti e proposte contenenti delle indicazioni sulle quali ci si potrebbe andare a focalizzare per produrre una modifica che effettivamente sia confrontabile o addirittura migliore di ciò che è già disponibile.

Indice

Elenco delle tabelle

Elenco delle figure

Capitolo 1

Concetti fondamentali

Un obiettivo fondamentale della visualizzazione e di chi pubblica grafici di dati è quello di produrre immagini che contengano dati che favoriscano l'analisi visiva, l'esplorazione e la scoperta di nuove informazioni. In questo contesto, un ruolo importante è ricoperto dalla percezione visiva. Come una persona vede e percepisce i dettagli in un'immagine può direttamente impattare sull'efficacia dell'osservazione. Una conoscenza di questo aspetto può migliorare significativamente la qualità e la quantità delle informazioni che vengono mostrate sull'immagine. In più, ciò che noi esseri umani vediamo con i nostri occhi è fortemente influenzato dalla direzione in cui stiamo guardando e verso quale zona la nostra attenzione è focalizzata. Un ulteriore aspetto da considerare è che cosa ci si aspetta di trovare in un'immagine prima ancora di andare ad osservarla, che può incidere e non poco sui risultati. Dopo numerosi studi e anni di ricerca si è notato che in ogni momento la visione dettagliata per forme e colori è possibile solo in una piccola porzione del campo visivo. In pratica per riconoscere queste due caratteristiche, l'occhio umano scansiona piccole porzioni alla volta dell'immagine che si sta fissando. Un'importante scoperta è stata poi quella di identificare un insieme di caratteristiche visive che possono essere individuate rapidamente dall'apparato visivo in una prima osservazione che è definita di *basso livello*. Queste sono state chiamate *preattentive* dato che la loro identificazione avviene prima ancora che l'attenzione si concentri sull'immagine. Tipicamente si considera *preattentive* tutto ciò che si verifica nei primi 200-250 ms. Alcune proprietà degli elementi presenti in un'immagine che sono classificate in questa maniera sono ad esempio: l'orientazione, la lunghezza, la dimensione, il colore, le ombre e alcune altre caratteristiche degli oggetti elencate nel documento [1]. E' interessante sottolineare come se queste caratteristiche vengano combinate nella stessa immagine, potrebbero non più essere identificabili in maniera rapida rispetto a quanto avviene quando si presentano da sole. Queste proprietà *preattentive* sono state usate in esperimenti per i seguenti scopi:

- *target detection*, in cui gli osservatori sono chiamati a ricercare un elemento che presenta un'unica caratteristica visiva;

- *boundary detection*, in cui gli osservatori devono dividere in due gruppi gli oggetti che sono caratterizzati da almeno una caratteristica in comune tra i due gruppi;

- *region tracking*, dove lo scopo è quello di tenere traccia di un oggetto che si muove nello spazio e nel tempo;

- *counting and estimation*, dove l'obiettivo è riuscire a contare il numero di elementi che presentano un'unica caratteristica visiva.

Sono state formulate numerose teorie che cercano di spiegare come il processo *preattentive* influenzi il sistema visivo .

Treisman [16][17] ha proposto un modello per l'attenzione umana nel basso livello dell'osservazione [1]. Questo è composto da un insieme di *feature maps* (ognuna legata ad una precisa caratteristica visiva ed agenti in parallelo) e una *master map of locations* (che invece introduce il concetto di posizione sull'immagine). La psicologa ha in più affermato che l'ammontare della differenza (cioè quanto sono diversi) tra il target e i distrattori presenti sull'immagine influisce sul tempo di ricerca.

Julész ha teorizzato come la prima parte della visione vada ad individuare tre categorie di proprietà visive chiamate *textons* [1][18]: le intersezioni tra le linee; i terminatori delle linee; le forme (linee, rettangoli, ellissi . . .) con specifica tonalità, orientazione e così via. La psicologa ha inoltre dichiarato che a suo parere solo una differenza tra vari *textons* possa essere notata nella fase *preattentive*.

Duncan e Humphreys [19] hanno asserito che il tempo di ricerca di un target è basato su due criteri: la similarità T-N e la similarità N-N [1]. La prima è la somiglianza tra gli obiettivi (*targets*) e i non obiettivi (*nontargets*), mentre la seconda è la similarità che è presente solo tra i *nontargets*. La variazione di una delle due influisce anche sull'altra, però in generale quando la similarità T-N aumenta allora il tempo di ricerca diminuisce e la stessa cosa accade quando quella N-N decresce. In più, i due ricercatori hanno proposto una teoria sulla selezione visiva organizzata in tre step: il campo visivo è segmentato in parallelo; l'accesso alla memoria visiva a breve termine è una risorsa limitata; una scarsa corrispondenza tra una zona e uno schema di ricerca di alcune proprietà visive porta ad automatiche esclusioni di regioni che sono legate ad una già esclusa.

Più recentemente Wolfe ha invece designato l'idea della ricerca guidata [20] ed è la prima volta in cui gli obiettivi dell'osservatore sono stati inseriti in un modello sulla ricerca visiva [1]. Egli ha ipotizzato che una *activation map* basata sull'approccio *bottom-up* (ciò che emerge automaticamente dall'immagine) e su

quello *top-down* (quello che è legato alla consapevolezza di chi guarda) si costruisse durante la visualizzazione, con dei picchi di intensità dove l'attenzione ricade maggiormente. Come Triesman, anche Wolfe è partito dall'idea di base secondo la quale l'immagine venga divisa in *feature maps* che poi ricombinate insieme generano i picchi sulla *activation map*.

In seguito, Huang ha presentato un nuovo modello di visione a basso livello concentrandosi sul cercare di capire per quale motivo spesso sull'immagine ci si concentra su degli elementi che in realtà non sono rilevanti con ciò che è realmente di interesse [1][21]. La sua idea si basa sul fatto che la ricerca visiva sia divisa in due parti: *selezione* e *accesso*. Durante la prima si selezionano un insieme di oggetti dalla figura, mentre nella seconda si determina quale proprietà degli oggetti selezionati possa essere appresa da un osservatore. Huang ha continuato affermando che il sistema visivo può suddividere l'immagine tra gli oggetti selezionati e quelli esclusi, per poi accedere a certe proprietà solo per i selezionati. Queste sono le fondamenta che stanno dietro alla *boolean map* [9][22].

Un aspetto rilevante che è di base per la visione di basso livello è la capacità di generare un rapido riepilogo di come le caratteristiche visive più semplici siano distribuite tra i campi visivi. Per la prima volta questo concetto è stato riportato da Ariely e prende il nome di *ensemble coding* [1][15].

Una delle più importanti considerazioni per un designer di informazioni visive è decidere come presentarle senza creare troppa confusione. E' importante sapere come per certi scopi, alcune caratteristiche visive possano essere più salienti, cioè più rilevanti, di altre [10]. Alla luce di questo è normale asserire che i dati più importanti che si vogliono mostrare dovrebbero essere associati alle caratteristiche visive che sono più salienti in quel determinato contesto.

Passiamo adesso ad alcune brevi informazioni che sono invece più legate all'eye tracking. Wolfe ha condotto degli studi per determinare se mostrare all'osservatore un'immagine (simile a quella dell'esperimento) prima della ricerca accresce l'abilità di chi guarda nel ricercare gli obiettivi. Intuitivamente ci si potrebbe attendere che la risposta sia affermativa, invece gli studi hanno dimostrato come questo non sia vero e per questo si parla di *postattentive amnesia* [23]. Scoperte più recenti hanno invece mostrato come la memoria pregressa può portare benefici nella ricerca visiva. Se ad esempio un sottoinsieme dell'immagine è ripetuto in prove successive, l'osservatore è in grado di ricercare l'obiettivo in maniera più rapida rispetto a quando l'obiettivo è posizionato in zone che non sono ancora mai state mostrate in precedenza [1].

In conclusione a questa prima parte introduttiva, ricordiamo che la visione è il senso dominante per gli esseri umani [25]. Esistono ricerche che testimoniano come più della metà del cervello è coinvolto nel processare informazioni visive. Per questo motivo comprendere l'attenzione visiva è importante sia nella visualizzazione che nella grafica. Essere in grado di tracciare l'attenzione dell'occhio umano può essere

un vantaggio per predirre dove un osservatore guarderà ed è uno degli aspetti centrali di questo lavoro. Essere in grado di seguire l'occhio nel suo movimento e raccogliere dei dati su cui lavorare permette di modellare le parti dell'immagine in maniera differente in base a quanta attenzione ci si aspetta che ogni singola zona riceverà.

Capitolo 2

Algoritmi predittivi

2.1 Algoritmo di Itti

Presentiamo adesso un modello relativo all'attenzione visiva che si ispira al comportamento dell'architettura neurale proposta dai primi sistemi visivi. Itti ha basato il suo studio della salienza sul movimento dell'occhio secondo la teoria delle *feature maps* proposta da Triesman [1][16][17]. Più caratteristiche grafiche vengono combinate in un'unica *saliceny map* con scale diverse e successivamente un *neural network* dinamico seleziona le zone in ordine di salienza decrescente.

Il modello (figura 2.1 [2]) è costruito partendo dalla teoria dell'integrazione delle caratteristiche, la quale descrive le strategie e le metodologie adottate nella ricerca visiva umana. L'input visivo, cioè l'immagine che l'osservatore sta fissando, è come prima cosa scomposto in tre insiemi di *feature maps*. Zone differenti dell'immagine competono per la salienza e solo quelle che si distinguono (nel senso che prevalgono e dopo capiremo in che senso) rispetto alle altre sopravvivono ad una prima scrematura. Tutte le mappe confluiscono poi in un'unica *saliency map* mediante delle semplici combinazioni lineari.

Entrando più nel vivo di ciò che accade, data in ingresso un'immagine a colori, vengono generati nove valori di scala utilizzando la piramide gaussiana [24] che progressivamente sottocampiona sempre di più l'immagine. Ogni proprietà visiva è poi calcolata con un insieme di operazioni lineari che legano il centro dell'immagine con ciò che c'è intorno. Per quanto riguarda i colori, sull'immagine in ingresso vengono utilizzati i canali RGB (*red*, *green* e *blue*) che forniscono un'intensità della tonalità dell'immagine come una media tra loro tre. Questa valore di intensità è usato per creare le piramidi gaussiane con i rispettivi valori di scala.

Il primo insieme di *feature maps* è costruito concentrandosi sul contrasto e quindi dove c'è una discrepanza maggiore tra la luminosità nel centro e sul bordo. Un secondo insieme di feature maps è invece realizzato partendo dai canali dei

colori. Infine l'ultimo gruppo è calcolato sulla base dell'orientazione usando le piramidi orientate di Gabor, nelle quali si combinano sia i valori di scala e sia alcuni valori per gli angoli di rotazione. Alla fine di questa prima parte sono presenti in tutto 42 *feature maps*: 6 per l'intensità, 12 per i colori e 24 per l'orientazione.

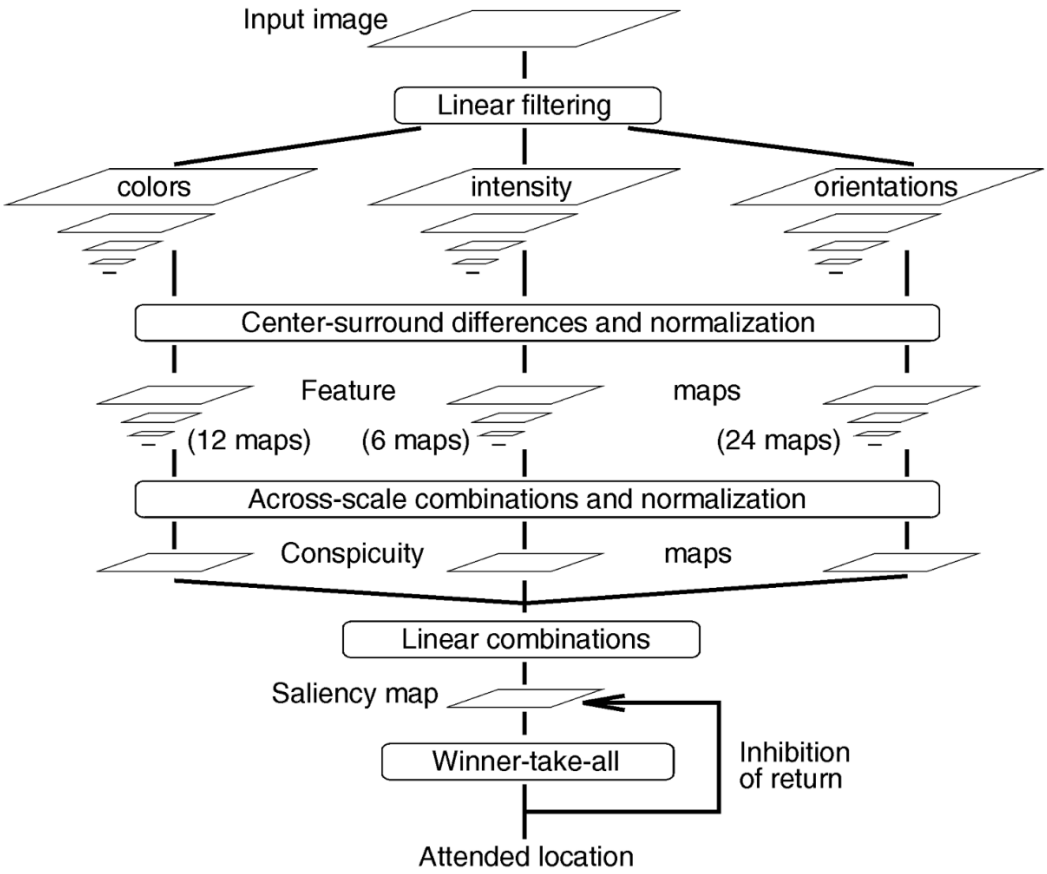

Figura 2.1. Schema algoritmo di Itti

Il passo successivo è quello di costruire la mappa di salienza e quindi di associare ad ogni zona dell'immagine un valore scalare che rappresenti quanto questa è potenzialmente rilevante. Come prima cosa viene effettuata una combinazione delle *feature maps* che è un input alla mappa di salienza (modellata come un *neural network* dinamico). Una delle difficoltà nel combinare insieme questi elementi è che alcune delle mappe sono difficili da confrontare tra di loro e gli elementi salienti appaiono rilevanti solo in poche di queste, quindi rischiano di perdersi una volta che vengono combinate tutte insieme. Per ovviare a questo problema è stato inserito un operatore di normalizzazione per ogni caratteristica e tre sono le mappe generali

calcolate (chiamate *conspicuity maps*): una per l'intensità, una per il colore e una per l'orientazione. Queste sono poi a loro volta normalizzate e ne viene fatta la media per ottenere la *saliency map* finale. In ogni istante, il valore massimo sulla mappa di salienza identifica la regione dell'immagine più saliente verso la quale l'attenzione di chi guarda dovrebbe essere diretta.

Il modello proposto ha un'architettura relativamente semplice, ma nonostante questo è in grado di fornire buone prestazioni con anche scene complesse riferite a situazioni naturali [4] (ad esempio fotografie di paesaggi o luoghi reali) e questo rinforza l'idea che un'unica mappa di salienza che riceve input dalle prime elaborazioni del processo visivo sia sufficiente per il processo *bottom up*. Da un punto di vista computazionale la forza di questo approccio è la parallelizzazione delle varie caratteristiche visive. Un difetto è invece la difficoltà nel rilevare come salienti degli elementi che presentano una combinazione di più caratteristiche visive. Una componente essenziale del modello è la normalizzazione, la quale fornisce un meccanismo generico per calcolare la salienza in qualunque situazione.

Altri modelli di questo tipo che verranno brevemente citati nella sezione successiva sono il BMS (*Boolean Map Based Saliency model*) e l'eDN (*Ensembles of Deep Networks Model*) che in linea generale si basano sugli stessi principi, con qualche piccola variazione.

I grafici di business, o in generale i grafici di dati (che sono quelli che interessano a noi), non hanno però le stesse caratteristiche delle immagini che ritraggono situazioni del mondo reale [7][11]. Nella sezione successiva vengono evidenziate le problematiche e viene presentata la soluzione proposta da Matzen che abbiamo adottato per i nostri scopi durante questo lavoro di tesi.

2.2 Algoritmo di Matzen

I modelli che determinano la saliency visiva (ad esempio Itti [2]) sono tipicamente basati sulle proprietà della corteccia visiva umana [26] e predicono quali aree di un'immagine hanno delle caratteristiche visive tali da attirare l'attenzione dell'osservatore. Questi modelli tipicamente possono prevedere abbastanza bene dove gli osservatori guarderanno in immagini di tipo naturale (quindi fotografie di situazioni reali), ma le loro prestazioni calano nell'ambito di visualizzazioni astratte (quindi per quanto riguarda ad esempio grafici di dati [13]). In questa sezione discuteremo i motivi di questo calo di performance ed è presentata una soluzione, ovvero l'algoritmo di Matzen [3].

Nel contesto di grafici di dati, le caratteristiche visive e la loro disposizione sull'immagine sono scelte per un particolare obiettivo [6]. Chi crea i grafici, organizza l'immagine per comunicare informazioni in maniera tale che chi sta guardando possa trovare ciò che è di interesse in maniera agile. Oltretutto, a differenza delle

immagini naturali, i grafici di business sono tipicamente di origine digitale, quindi generati da computer ed è più facile avere alcune zone isolate della figura come ad esempio degli specifici gruppi di dati oppure delle porzioni di testo.

I modelli sulla *saliency* visiva che inizialmente Matzen ha preso come riferimento sono stati Itti, eDN e BMS [27]. Tutti seguono un approccio comune in cui il primo passo è calcolare delle mappe di interesse a varie risoluzioni e successivamente combinarle insieme per costituire la mappa di salienza finale. Questa strategia funziona bene per immagini naturali, ma nella visualizzazione astratta di dati le proprietà spaziali sono diverse: alcuni elementi (linee o porzioni di testo) sono tipicamente più piccoli e nel processo di sottocampionamento si rischia di perderne alcuni dettagli. Ad esempio, con ciascuno dei modelli appena citati, il testo diventa sfocato e perde il suo significato nel corso delle analisi. Un'altra problematica di questi algoritmi già esistenti è che spesso come primo step viene effettuato un riscalamento dell'immagine ad una dimensione standard e questo può essere grave perchè porta gli elementi grafici a deformarsi e quindi a perdere le loro proprietà.

Un aspetto incisivo su cui concentrarsi è quello dei colori [28]. La percezione umana del colore è differente tra colori su carta oppure su un display elettronico e quindi la modifica proposta è stata quella di lavorare con lo spazio dei colori *CIE LAB* che utilizza tre canali: due di colore e uno per la luminosità dell'immagine.

Una differenza cruciale tra grafici di dati e le immagini naturali è costituita sicuramente dalla presenza di grandi spazi bianchi (tipicamente) in cui sostanzialmente non ci sono informazioni da rilevare. Il particolare però più rilevante di tutti è senza dubbio quello del testo [30]. Questo in generale non è presente (oppure se sì in quantità limitata) nelle immagini naturali, mentre costituisce un elemento importante nei grafici di dati per trasmettere informazioni. Da ricerche che sono state effettuate, più del 60% [29] dell'attenzione ricade esclusivamente su parti scritte, a discapito di zone contenenti parti puramente grafiche. Dato che, come già scritto in precedenza, nei primi modelli di *saliency* citati il testo nel corso dell'analisi diventa sfocato, è stato necessario introdurre una parte dell'algoritmo che si concentrasse esclusivamente su di esso. Il nuovo modello di saliency presentato per i grafici di dati, chiamato modello DVS (*Data Visualization Saliency* disponibile a questo link), è costituito da due parti principali: la prima è una versione modificata del modello di Itti e la seconda si occupa di riconoscere il testo.

Come primo riferimento di partenza, è stato preso Itti perchè sul dataset MASSVIS sul quale ci si è concentrati (che è descritto nella sezione successiva) è il modello sulla saliency visiva con le migliori prestazioni [8] (in relazione ad otto metriche, anche queste descritte nella prossima parte). La modifica effettuata riguarda i colori. Come raccontato in 2.1, Itti originale utilizza il riconoscimento dei colori tramite i canali RGB, ma dato che non sono più trattate immagini naturali e i colori possono essere scelti liberamente da chi crea i grafici, è bene cambiare la metodologia ed è stata adottata la rappresentazione delle immagini in input

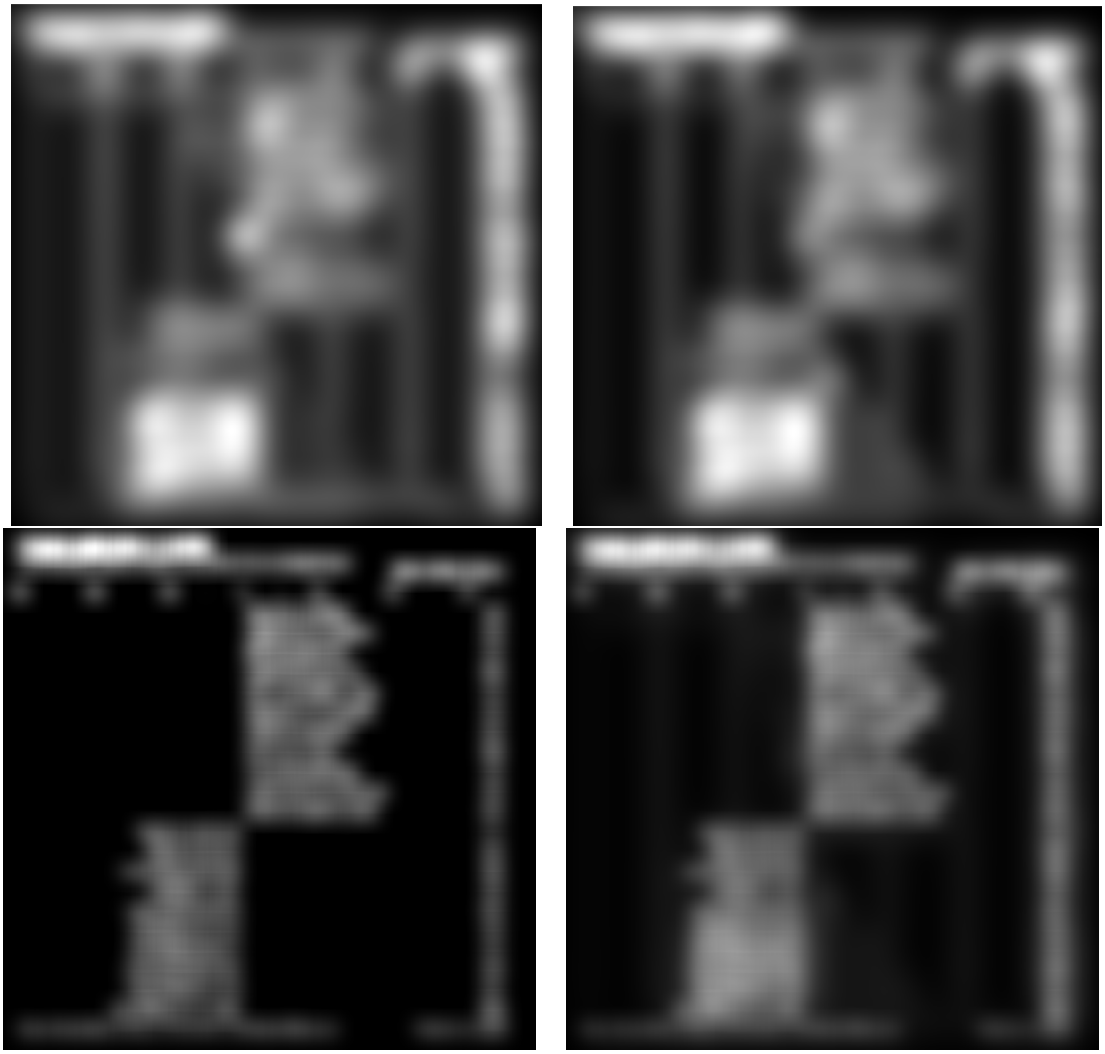

Figura 2.2. Analisi algoritmo Matzen

tramite lo spazio dei colori *CIE LAB*. Nella parte superiore della figura 2.2 sono messi a confronto i risultati prodotti dalle analisi di Itti originale (a sinistra) e Itti modificato (a destra) sulla prima immagine in figura 2.3 del dataset di MASSVIS. Come si può notare, il cambiamento a livello visivo non è eclatante, infatti la modifica è maggiormente rilevante a livello di prestazioni numeriche delle metriche su questo tipo di immagini, in particolare quando Itti modificato (che a livello grafico produce una mappa meno sfocata rispetto all'originale) viene combinato insieme all'analisi del testo.

Come già accennato in precedenza, si è scoperto che nei grafici di dati gli osservatori dedicano la maggior parte dell'attenzione nei confronti del testo. Spesso

queste porzioni tendono ad avere un elevato contrasto, ad essere piccole e in più l'elevata frequenza in queste zone tende ad essere persa quando si effettuano dei riscalamenti. A tal proposito Matzen ha dedicato una porzione del modello che si occupa di riconoscere esclusivamente il testo per poi poter essere combinata con Itti modificato. Il metodo utilizzato per questo scopo è una combinazione di varie tecniche classiche di riconoscimento del testo già esistenti. A differenza di queste però, non viene prodotto un output binario (testo c'è/testo non c'è) bensì viene generata una distribuzione continua di probabilità che può essere incorporata ad una *saliency map*. Più nel dettaglio l'approccio utilizzato è stato l'MSER [31] (*Maximally Stable Extremal Regions*) con il quale si cercano delle regioni che sono possibili candidate a contenere del testo (regioni di pixel omogenee e connesse), per poi applicare su queste alcuni indicatori che fanno da filtro e permettono di escludere chi non ne contiene. In basso a sinistra in figura 2.2 è presente il risultato dell'analisi sul testo sempre della prima immagine in figura 2.3 e possiamo osservare come tutto il testo venga correttamente identificato.

Infine quello che Matzen ha eseguito è stata una combinazione lineare tra Itti modificato e l'analisi sul testo per generare la sua mappa di salienza scegliendo accuratamente i pesi con cui unire le due parti confrontando i risultati con quelli di *MASSVIS* [32]. Nel complesso questa produce sulle metriche con cui vengono misurate le prestazioni risultati migliori rispetto ai precedenti algoritmi. Nel documento [3] si trovano le tabelle numeriche riassuntive che dimostrano la superiorità dell'algoritmo di Matzen per quanto riguarda il dataset fornito da *MASSVIS*. Nelle sezioni seguenti andremo anche noi a verificare queste informazioni per cercare di capire se è presente qualche difetto e se si è in grado di proporre o di dare dei suggerimenti per un algoritmo migliore o competitivo. In basso a destra in figura 2.2 è mostrata la combinazione tra Itti modificato (in alto a destra) e l'analisi sul testo (in basso a sinistra) della prima immagine in figura 2.3. Come emergerà anche successivamente dall'esperimento da noi creato, quando il testo è presente in maniera abbondante sull'immagine attira quasi tutta l'attenzione su di sè e ne abbiamo avuta una prima conferma. Ad esempio, per la mappa di salienza combinata appena citata, ciò di cui rimane traccia è per lo più del testo e la parte grafica (quindi le barre) praticamente non è evidenziata.

2.3 Dataset e verifiche

Come già espresso nella sezione precedente, il lavoro di Matzen si è concentrato su analizzare i dati forniti da *MASSVIS*. In questa parte entriamo più nel dettaglio dei risultati da lei ottenuti e mostriamo ciò che abbiamo replicato. *MASSVIS* è un dataset per la visualizzazione di immagini raffiguranti grafici di dati e le cui

fonti sono: i dati messi a disposizione dai governi, i grafici informativi, le notizie e la scienza.

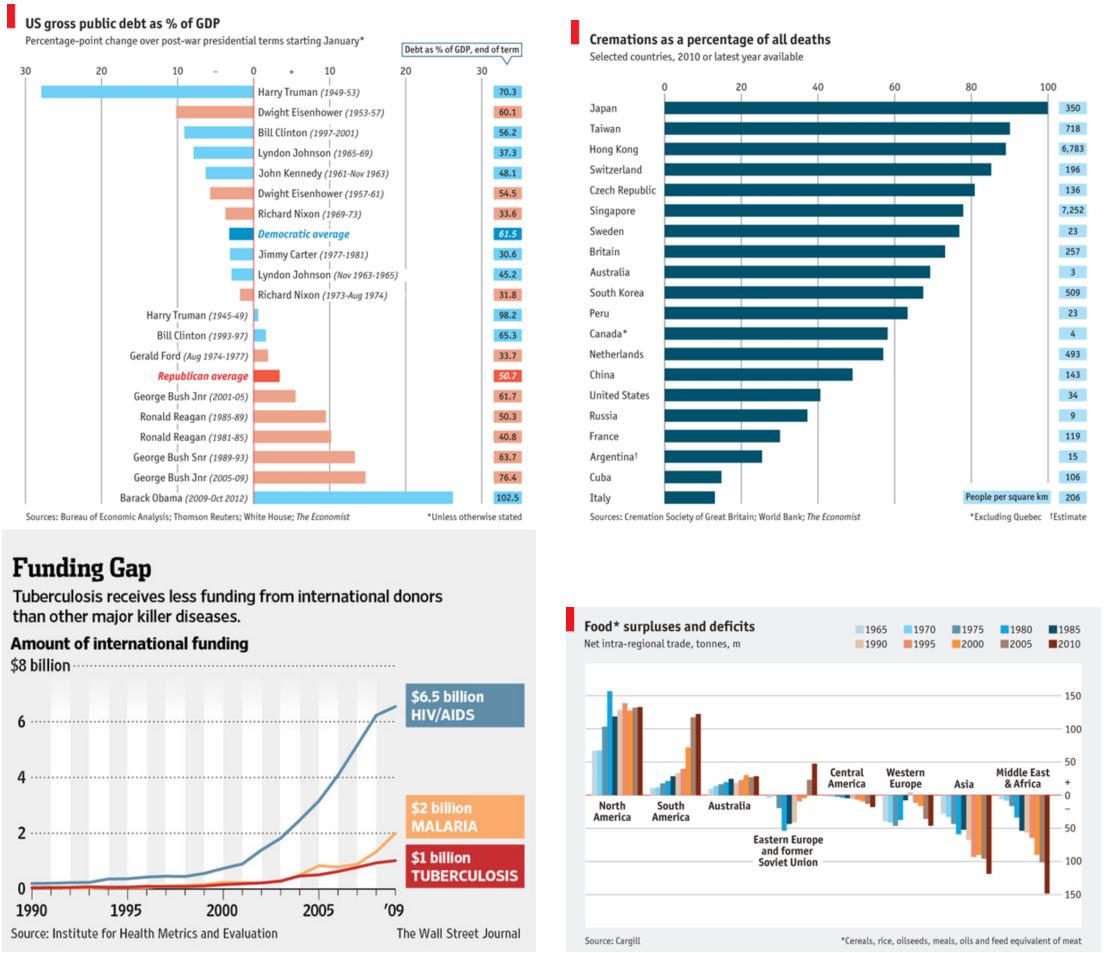

Figura 2.3. Dataset MASSVIS

Nel link al dataset precedente (indicato in blu) è presente una sezione che chiamata *Eye-movement Data* in cui ci viene descritto l'esperimento di eye tracking che è stato eseguito per raccogliere i dati. Il dataset è composto da 393 immagini (di vario genere) che sono state mostrate per 10 secondi ciascuna a 33 osservatori dei quali si è tenuta traccia del movimento dell'occhio. Ogni immagine è stata vista da almeno 16 osservatori.

In figura 2.3 sono rappresentati quattro esempi di ciò che *MASSVIS* contiene per avere un'idea di quali tipologie di grafici si sta parlando. Da sinistra verso destra e dall'alto verso il basso i nomi delle immagini sono rispettivamente: *economist_daily_chart_4* , *economist_daily_chart_5*, *economist_daily_chart_103*

e *wsj612*. Precisiamo che delle 393 immagini presenti nel dataset, ne abbiamo

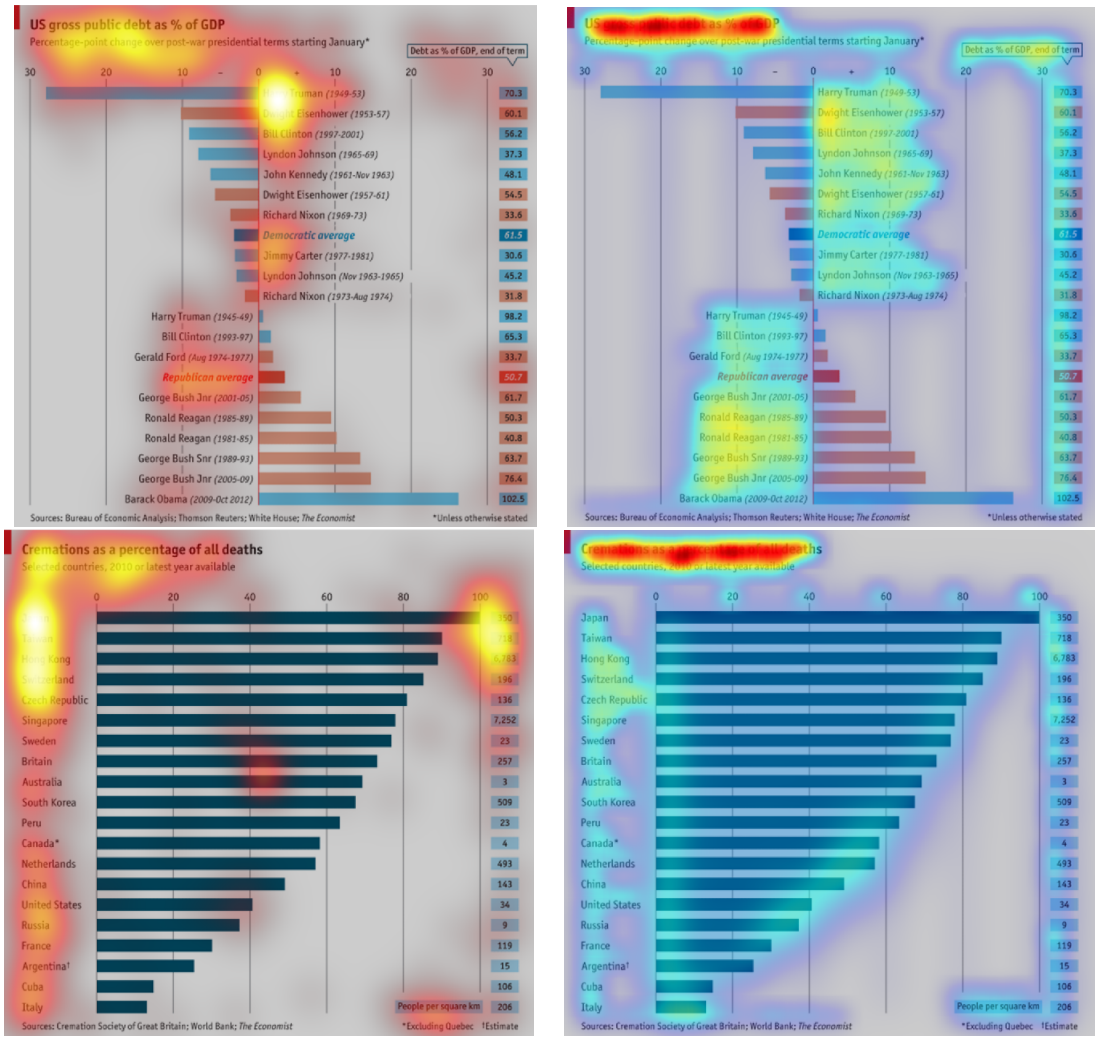

Figura 2.4. Confronto punti di fissazione e Matzen

conservate 110 per effettuare le nostre analisi, in quanto alcune avevano sfondi troppo elaborati oppure non erano dei veri e propri grafici di dati e quindi non erano di nostro interesse per questo lavoro. Nella sezione 4.3 di questo documento viene descritto tutto il materiale che abbiamo elaborato e sono presenti anche le informazioni relative a quali parti di *MASSVIS* abbiamo conservato.

Scaricando il materiale *Data* e il codice Matlab *Code* dal sito di *MASSVIS* abbiamo potuto riprodurre le mappe di calore su ognuna delle immagini e confrontarle successivamente con le mappe di saliency che l'algoritmo di Matzen genera. Nelle figure 2.4 e 2.5 sono mostrate: a sinistra le immagini originali di *MASSVIS* alle

quali è stata sovrapposta la mappa di calore in base ai dati dell'eye tracking e a destra le mappe che vengono invece prodotte da Matzen per quelle immagini.

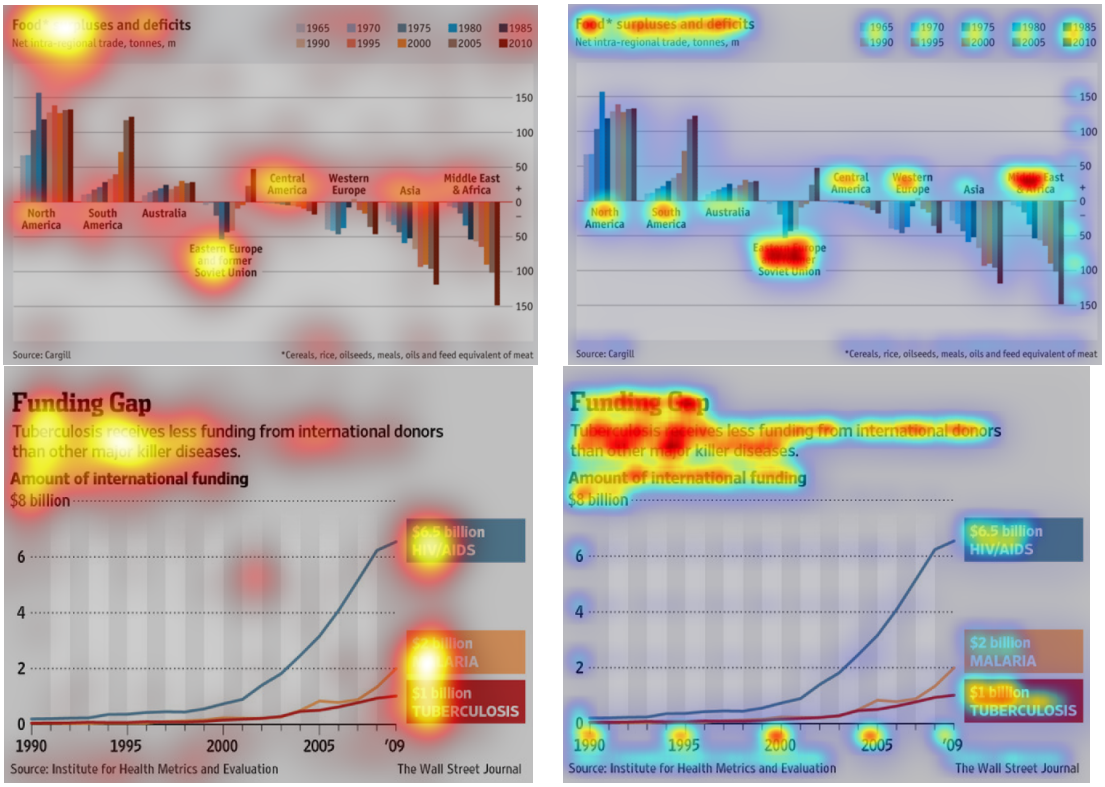

Figura 2.5. Confronto punti di fissazione e Matzen

Entreremo meglio nel merito di quanto graficamente viene mostrato, ma a primo impatto possiamo notare come la parte di testo presente sull'immagine (titolo, sottotiolo, legenda oppure etichetta dei dati) attiri la maggior parte dell'attenzione su di essa mentre le parti puramente grafiche sembrano essere un pò trascurate e lasciate in secondo piano.

Valutare le prestazioni di Matzen solamente da un punto di vista grafico non è però sufficiente. E' bene quantificare le sue prestazioni anche a livello numerico in relazione a quanto la mappa di salienza che produce si discosta da quella vera. A tal proposito, per questo tipo di calcoli sono disponibili otto metriche [5]: le prime tre si occupano di quantificare quanto bene la mappa di salienza di Matzen predice le zone in cui l'occhio umano si è realmente concentrato; le successive quattro eseguono un confronto tra la distribuzione delle fissazioni sull'immagine e la distribuzione della salienza nella *saliency map* generata dall'algoritmo; l'ultima esegue un'analisi un pò diversa descritta di seguito. In dettaglio le metriche sono:

- *AUC-Borji*: questa metrica misura quanto la mappa di salienza generata su un'immagine predice correttamente le regioni di fissazione dell'occhio umano;
- *AUC-Judd*: il suo compito è lo stesso della precedente, ma cambia come sono gestite le curve ROC (*Receiver operating characteristic*) e la gestione dei falsi positivi;
- *AUC-Shuffled*: anche lei ha lo stesso ruolo delle due metriche precedenti, ma cambia ancora una volta come sono gestite le curve ROC (*Receiver operating characteristic*) e la gestione dei falsi positivi;
- *CC*: questa metrica calcola il coefficiente di correlazione lineare secondo Pearson tra due differenti mappe di salienza, in cui falsi negativi e falsi positivi sono penalizzati equamente;
- *EMD*: questa metrica calcola il coefficiente denominato *emd_hat* descritto nel paper [12]. In particolare questo coefficiente rappresenta il costo di trasformare una mappa nell'altra, quindi più è alto e più significa che il risultato ottenuto è peggiore;
- *KL*: questa metrica trova la divergenza secondo Kullback e Leibler tra due differenti mappe di salienza quando sono trattate come distribuzioni. Fornisce un dato di quanta informazione viene persa quando la mappa di salienza dell'algoritmo è usata al posto di quella vera creata dai dati raccolti sugli osservatori. Come per *EMD*, anche in questo caso quando otteniamo un numero più alto significa che l'algoritmo ha performato peggio;
- *SIM*: questa metrica tratta le fissazioni e la mappa di salienza come istogrammi e calcola quanto si sovrappongono bene l'uno sull'altro;
- *NSS*: questa metrica trova il *saliency scanpath* normalizzato tra due differenti mappe di salienza. Standardizza la mappa di salienza dell'algoritmo e poi calcola la *saliency* media in posizioni predefinite.

Partendo dalle 110 immagini di *MASSVIS* che sono state conservate e servendoci delle metriche appena descritte mediante i loro rispettivi codici Matlab, abbiamo costruito la tabella 2.1. In essa sono riassunti i risultati su ognuna delle otto metriche calcolate per tre algoritmi: Itti originale, itti modificato e Matzen. In particolare la tabella è divisa in tre sezioni in maniera tale che siano confrontati tra di loro due algoritmi alla volta. Il raffronto tra un modello e l'altro è in termini di percentuale di vittorie su ognuna delle metriche. Prese in riferimento un'immagine e una metrica, vince l'algoritmo per il quale il valore della metrica associato alla mappa di salienza da lui prodotta è superiore. Rispettivamente nella prima sezione abbiamo le percentuali di vittorie tra Itti originale (*ITTI OR*) e Itti

modificato (*ITTI MOD*), nella seconda c'è il confronto tra Matzen (*MATZEN*) e Itti originale e nella terza abbiamo Matzen comparato con Itti modificato.

	AUC-Borji	AUC-Judd	AUC-S.	CC	EMD	KL	NSS	SIM
ITTI OR	40,91%	37,27%	48,18%	50,91%	54,55%	64,55%	40,91%	48,18%
ITTI MOD	59,09%	62,73%	51,82%	49,09%	45,45%	35,45%	59,09%	51,82%
PAREGGIO	0%	0%	0%	0%	0%	0%	0%	0%
	AUC-Borji	AUC-Judd	AUC-S.	CC	EMD	KL	NSS	SIM
MATZEN	81,82%	97,27%	75,45%	90,90%	3,64%	0,91%	98,18%	85,45%
ITTI OR	18,18%	2,73%	24,55%	9,10%	96,26%	99,09%	1,82%	13,64%
PAREGGIO	0%	0%	0%	0%	0%	0%	0%	0,91%
	AUC-Borji	AUC-Judd	AUC-S.	CC	EMD	KL	NSS	SIM
MATZEN	82,73%	95,45%	80%	88,18%	3,64%	4,55%	97,27%	86,36%
ITTI MOD	17,27%	4,55%	20%	11,82%	96,26%	95,45%	2,73%	13,64%
PAREGGIO	0%	0%	0%	0%	0%	0%	0%	0%

Tabella 2.1. Risultati metriche globali MASSVIS

E' importante ricordarsi di tenere in considerazione il fatto che per le metriche *EMD* e *KL* i ragionamenti sono invertiti perchè vanno a misurare una dispersione rispetto alla vera mappa di salienza. Quindi, relativamente a queste due, chi tra gli algoritmi ha la percentuale di vittorie più alta significa che ha in realtà performato peggio.

Analizzando le percentuali in tabella abbiamo la conferma che Matzen sia effettivamente, almeno su queste immagini, l'algoritmo migliore tra i tre di cui stiamo discutendo (come raccontato in [3]). Rispetto ad Itti originale è superiore in tutte le metriche in almeno il 75% delle immagini, mentre in confronto ad Itti modificato è migliore in almeno l'80% delle immagini in ogni metrica.

Nelle prossime sezioni cercheremo di trovare dei difetti a questo algoritmo per capire se performa meglio degli altri due anche apportando delle piccole modifiche alle immagini di *MASSVIS* che abbiamo conservato.

2.4 Analisi semantica

In questa parte ci si focalizza sull'*analisi semantica*, ovvero sulla collazione dei poligoni forniti dal dataset sull'immagine e sulla distribuzione della *saliency* su di essi. Ogni poligono circonda una determinata porzione dell'immagine: il titolo, il sottotitolo, altre porzioni scritte sull'immagine, la legenda oppure le parti puramente grafiche (andando in alcune più nel dettaglio rispetto ad altre). Per essere più precisi il dataset *MASSVIS* mette a disposizione un file di testo per ogni immagine e questo contiene le informazioni sui punti che delimitano ogni poligono.

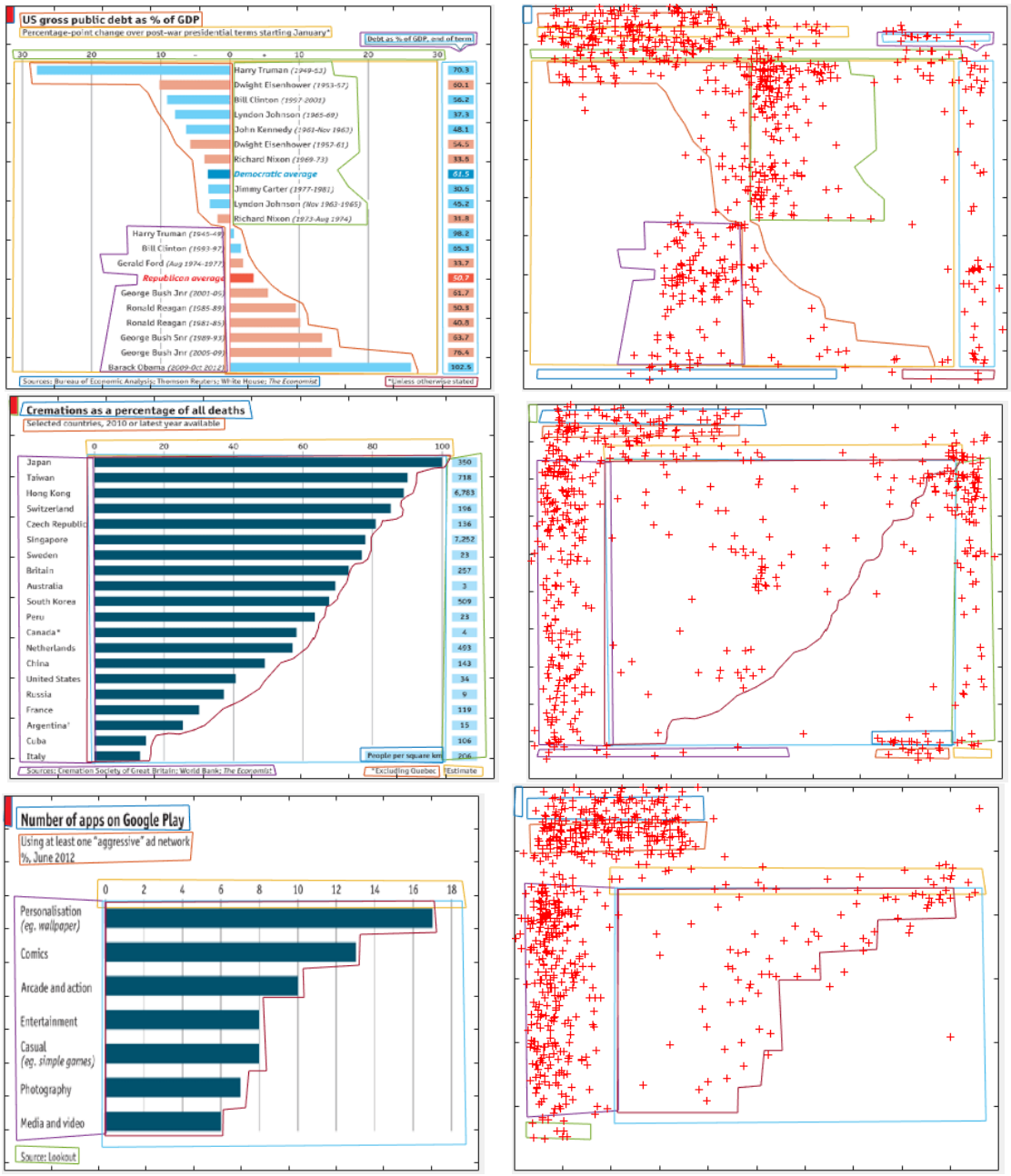

Figura 2.6. Analisi semantica

Lo scopo di questa analisi è come prima cosa mostrare dove sono collocati questi poligoni. In seguito, alla luce di quanto è emerso dall'analisi nella sezione

precedente, abbiamo rappresentato anche i punti di fissazione in una nuova immagine con sfondo bianco contenente solo i poligoni. In questa maniera siamo stati in grado di verificare anche graficamente che questi punti fossero maggiormente concentrati nelle zone contenenti del testo, dato che fino ad ora non avevamo avuto informazioni sui singoli punti di fissazione, ma avevamo a disposizione solo la mappa di calore dalla quale poter estrarre informazioni e commenti. In figura 2.6 sono riportati i risultati grafici di tre immagini contenute in MASSVIS, in particolare: *economist_daily_chart_4*, *economist_daily_chart_5* ed *economist_daily_chart_75*. Come si può notare, a sinistra abbiamo le immagini originali alle quali abbiamo sovrapposto i poligoni, mentre nella parte destra abbiamo i poligoni associati all'immagine con segnati anche i punti di fissazione. Possiamo confermare quindi che come era teorizzato e come avevamo osservato anche con le mappe di calore, la distribuzione dei punti di fissazione (e di conseguenza l'attenzione di chi guarda) è maggiormente concentrata nelle porzioni dell'immagine che contengono testo e non tanto sulle parti grafiche.

2.5 Testo

Alla luce del fatto che Matzen suddivida l'analisi in due parti per poi unire insieme i risultati e in base alle ricerche secondo le quali il testo è ciò che maggiormente attira l'attenzione dell'occhio umano, potrebbero sorgere spontanee alcune domande:

-"Non è che Matzen prevale sugli altri due algoritmi (Itti ed Itti Modificato) solamente perchè è presente del testo?"

-"Se dall'immagine tagliamo il titolo oppure altre parti di testo, come performa a livello di risultati rispetto agli altri due?"

-"Dopo aver ritagliato l'immagine escludendo parte del testo, quali sono le sue prestazioni grafiche, cioè la mappa di salienza, confrontandolo con i punti veri di fissazione degli osservatori, cioè la mappa di calore?"

Con lo scopo di cercare di rispondere alle precedenti domande abbiamo preso in osservazione 11 immagini di *MASSVIS* e abbiamo effettuato dei tagli andando ad escludere alcune delle parti di testo. Alcune le abbiamo tagliate anche più volte lasciando fuori prima solo il titolo, poi anche altre parti come ad esempio la legenda. Controllare la sezione 4.3 di questo documento relativa al materiale allegato per vedere e comprendere meglio come è organizzato tutto il materiale che è stato elaborato. Dopo aver effettuato i tagli abbiamo ripetuto tramite codice Matlab le stesse analisi che abbiamo già raccontato nelle parti precedenti di questo documento, cioè:

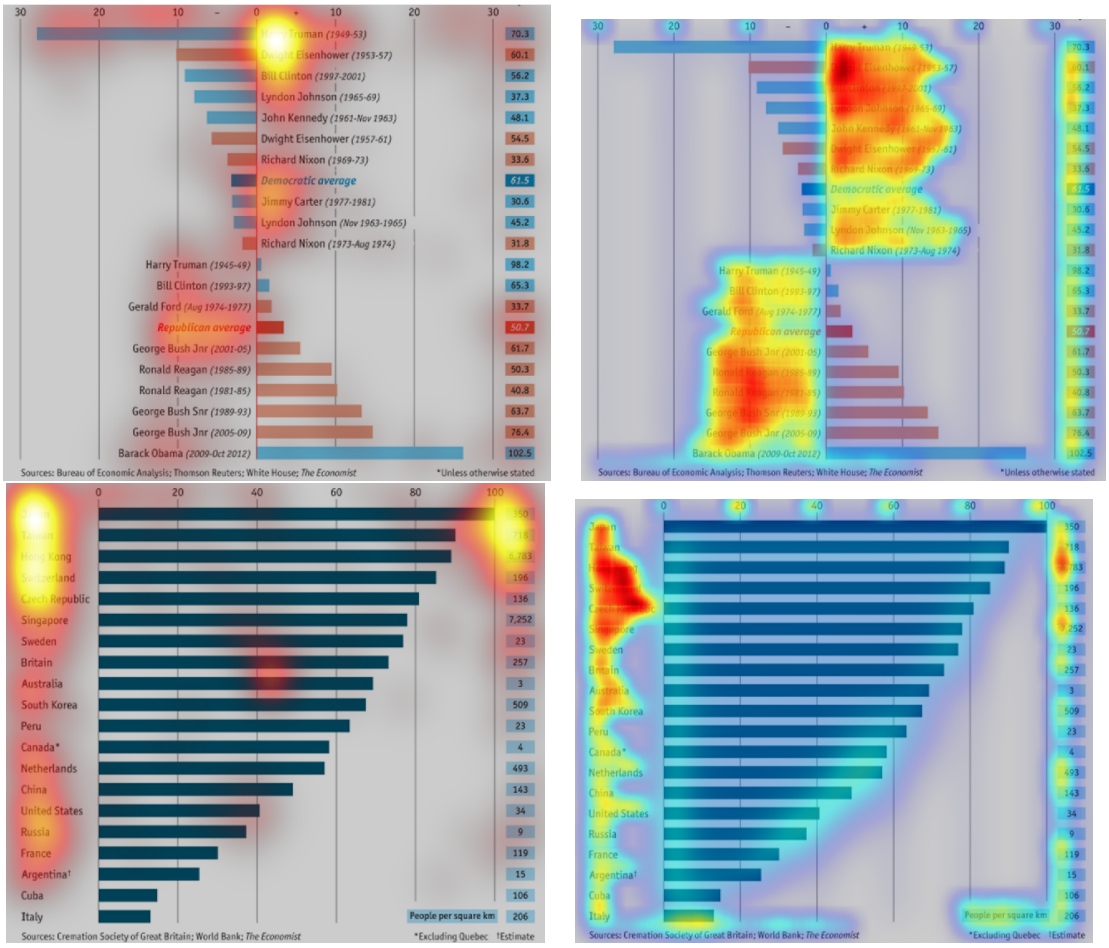

Figura 2.7. Analisi sul testo

- abbiamo riprodotto la mappa di calore relativa ai punti di fissazione degli osservatori (i punti che non rientrano più nella nuova immagine tagliata non sono stati presi in considerazione);

- abbiamo eseguito l'algoritmo di Matzen che genera la sua mappa di salienza di previsione dell'attenzione;

- abbiamo successivamente calcolato le metriche confrontando la mappa prodotta da itti, Itti modificato e Matzen con la mappa di calore vera.

In figura 2.7 sono contenuti due esempi di quello che abbiamo svolto su due delle immagini di *MASSVIS*: *economist_daily_chart_4* e *economist_daily_chart_5*. A sinistra sono contenute le immagini ritagliate con sovrapposta la mappa di calore

dei punti di fissazione, mentre a destra sono contenute le mappe di salienza generate da Matzen. In questi due casi particolari mostrati, nel ritaglio che abbiamo adoperato siamo andati ad escludere rispetto alle immagini originali solo il titolo.

Possiamo incominciare a notare come rispetto ai risultati grafici in figura 2.4 Matzen inizi ad andare un pò in confusione nel rappresentare la corretta porzione sulla quale con buona probabilità l'attenzione dell'occhio si andrà a concentrare. In particolare nell'immagine in alto possiamo notare come quando appena il titolo sia assente Matzen vada a scegliere in maniera quasi uniforme tutta la restante porzione di testo (cosa che logicamente non è corretta perchè viene ricoperta una superficie troppo estesa con la *saliency*). Similmente, ma in maniera meno accentuata, questo accade anche nell'immagine sottostante in cui comunque Matzen continua ad indicare in maniera corretta la porzione più saliente.

Quello che ora bisogna capire è se comunque, una volta che il titolo sparisce, l'algoritmo di Matzen, nonostante il difetto in molti casi di concentrarsi su tutto il testo a disposizione in maniera quasi uniforme, sia ancora l'algoritmo più efficiente rispetto a quello di Itti e a quello di Itti modificato per quanto riguarda il calcolo sulle metriche. A tal proposito nella tabella 2.2 sono presenti i risultati delle metriche confrontando due algoritmi alla volta come già fatto nella sezione precedente con la tabella 2.1, ma in questo caso solo per le immagini ritagliate. Ciò che emerge analizzando le percentuali di vittorie è che tagliando parte del

	AUC-Borji	AUC-Judd	AUC-S.	CC	EMD	KL	NSS	SIM
ITTI OR	20%	25%	50%	45%	40%	80%	25%	25%
ITTI MOD	75%	75%	50%	50%	60%	20%	75%	75%
PAREGGIO	5%	0%	0%	5%	0%	0%	0%	0%
	AUC-Borji	AUC-Judd	AUC-S.	CC	EMD	KL	NSS	SIM
MATZEN	85%	85%	80%	80%	25%	10%	95%	70%
ITTI OR	15%	15%	20%	20%	75%	90%	5%	30%
PAREGGIO	0%	0%	0%	0%	0%	0%	0%	0%
	AUC-Borji	AUC-Judd	AUC-S.	CC	EMD	KL	NSS	SIM
MATZEN	85%	85%	90%	80%	20%	10%	95%	65%
ITTI MOD	15%	10%	10%	15%	75%	85%	0%	30%
PAREGGIO	0%	5%	0%	5%	5%	5%	5%	5%

Tabella 2.2. Risultati metriche analisi sul testo

testo Matzen peggiora le sue percentuali di vittoria per quasi tutte le metriche rispetto agli altri due algoritmi (solo per *AUC-Borji* e *AUC-Shuffled* c'è in realtà un incremento), ma rimane comunque l'algoritmo di previsione migliore. Questo risultato è un'ulteriore conferma della solidità del modello.

Capitolo 3

Esperimento con l'eyetracker

3.1 Raccolta dei dati e dataset

Dato che le analisi svolte nel capitolo precedente sono legate ad un dataset già presente online e non controllato da noi, con immagini che spesso presentano troppe caratteristiche insieme, abbiamo pensato di crearne uno nuovo, con le caratteristiche che volevamo e che abbiamo deciso noi, con lo scopo di poter effettuare un nostro esperimento in maniera autonoma. Abbiamo dunque creato 30 immagini tutte legate ad una caratteristica di base: si tratta di grafici a barre, cioè di istogrammi, relativi all'epidemia di *COVID-19* e ai dati dei decessi in alcuni paesi del mondo (dati aggiornati a novembre 2020). Partendo da questo elemento abbiamo generato le immagini tramite python sfruttando le indicazioni a questo link. Tra un'immagine e l'altra sono presenti alcune variazioni che sono elencate qui di seguito:

- *barre orizzontali/verticali*: alcune immagini hanno le barre disposte orizzontalmente e alcune verticalmente;
- *titolo sì/no*: alcune immagini hanno il titolo in alto con il relativo sottotitolo e altre no;
- *titolo assi sì/no*: alcune immagini hanno il titolo sui due assi e altre no;
- *etichette barre sì/no*: alcune immagini hanno i numeri associati alla dimensione di ogni singola barra e altre no;
- *etichette assi sì/no*: alcune immagini contengono il valore numerico o l'etichetta di testo sugli assi e altre no;

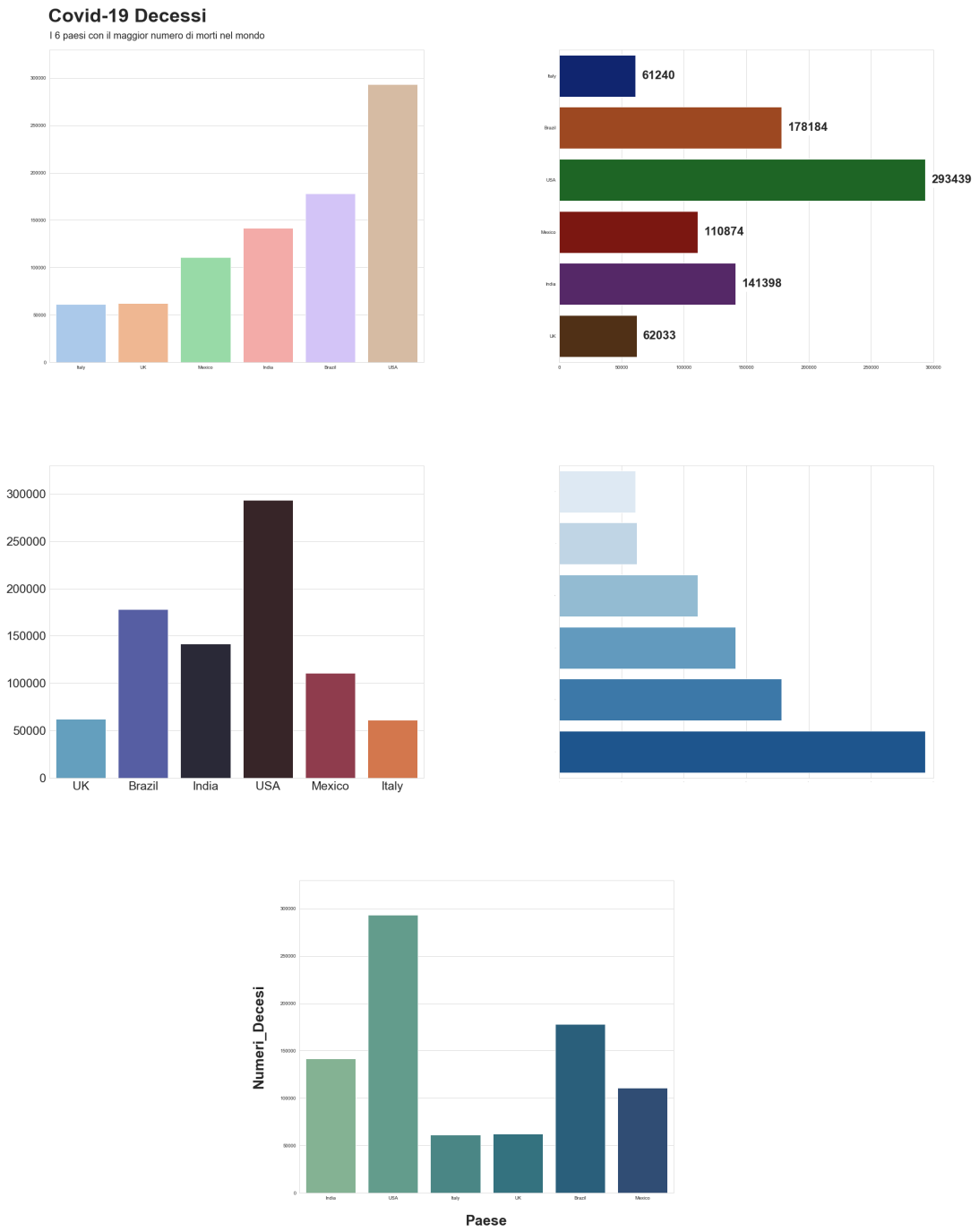

Figura 3.1. Dataset esperimento eye-tracker

- *etichette assi grandi/piccole*: in alcune tra le immagini che contengono l'etichetta sugli assi è stato utilizzato un carattere più piccolo e altre in altre un carattere un pò più grande;
- *colori accesi/tenui*: alcune immagini hanno le barre con colori accesi e altre invece le hanno con colori tenui;
- *colori diversi/scala unico colore*: alcune immagini contengono le barre che hanno colori diversi e invece in altre è presente la scala di un unico colore.

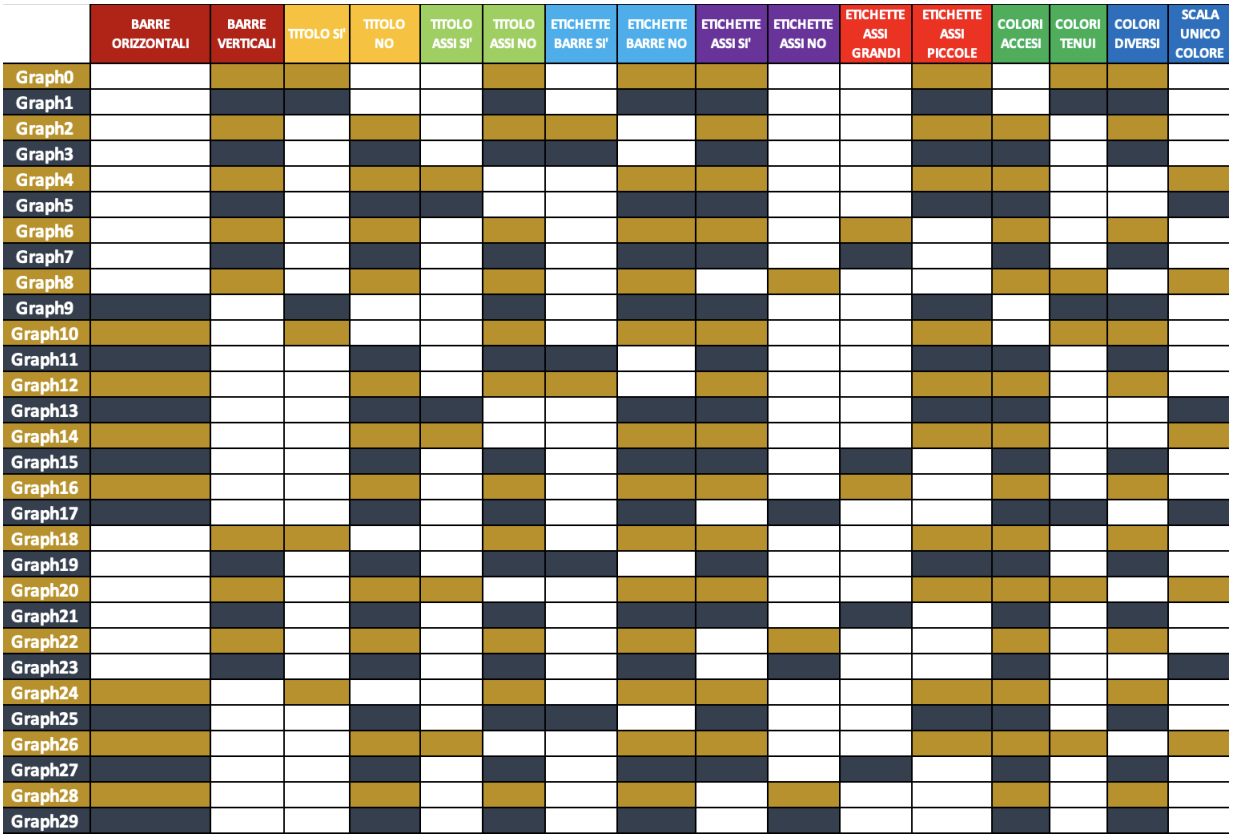

Figura 3.2. Distribuzione caratteristiche immagini dataset esperimento eye tracker

In figura 3.1 sono contenute 5 delle 30 immagini totali che rappresentano i prototipi di tutte le caratteristiche che ho elencato in precedenza. In 3.2 abbiamo invece un riassunto in formato di tabella che descrive come sono distribuite le caratteristiche tra le varie immagini del nostro dataset (sarà utile in seguito per commentare i risultati). Ricordo che nella sezione 4.3 viene descritto come è organizzato tutto il materiale elaborato e saranno anche presenti le analisi riguardanti il nostro esperimento.

La tabella 3.1 descrive come è distribuito il campione di osservatori al quale è stato sottoposto l'esperimento. In totale gli osservatori sono 62. Siamo consa-

	14/18	24/30	45/50
ETA'	51	3	8
	F	M	
SESSO	5	57	

Tabella 3.1. Campione esperimento eye tracker

pevoli del fatto che il campione non sia distribuito in maniera bilanciata (sia per quanto riguarda l'età e sia per quanto riguarda il sesso), però visto il periodo di emergenza sanitaria durante questi mesi in cui abbiamo portato avanti il lavoro di tesi non è stato possibile raccogliere in tempi brevi un insieme di osservatori con caratteristiche più varie e ci riteniamo fortunati già solo di poter aver effettuato l'esperimento.

Il Politecnico di Torino ci ha permesso di attrezzarci adeguatamente e siamo potuti entrare in possesso di un eye tracker con cui effettuare le varie prove: in particolare è stato adoperato il modello Tobii Pro Nano, mentre l'esperimento è stato svolto usando l'applicazione Open Sesame [14].

L'esperimento si è svolto come segue:

- inizialmente ogni osservatore è stato posizionato davanti al computer al quale è stato collegato l'eye tracker;
- in seguito è stata eseguita la calibrazione dell'eye tracker sull'osservatore per trovare la corretta distanza dallo schermo e l'altezza degli occhi rispetto allo strumento;
- dopodichè è iniziato il vero e proprio esperimento in cui ogni osservatore ha visualizzato ciascuna delle 30 immagini del dataset per 5 secondi in ordine casuale.

E' bene sottoineare che tra un'immagine e la successiva è stata inserita la pressione di un tasto della tastiera per permettere all'osservatore di riposizionarsi con lo sguardo nella parte centrale dello schermo. Questo ultimo passaggio è stato svolto con lo scopo di evitare che da un'immagine a quella successiva ci si portasse dietro un errore dovuto alla transizione tra una e l'altra senza la presenza di uno stacco di nessun tipo tra le due immagini.

3.2 Analisi sui dati raccolti

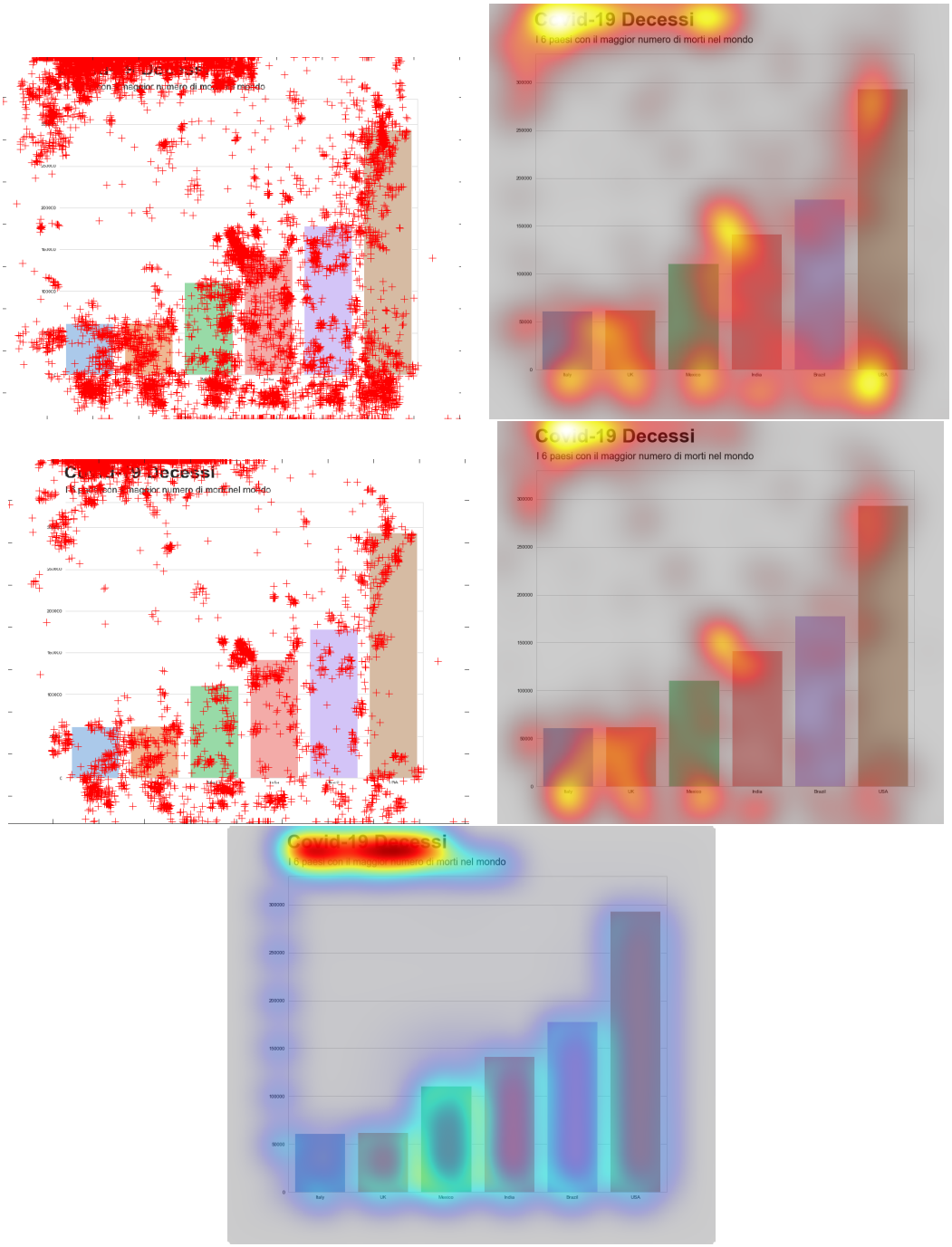

Figura 3.3. Punti di fissazione - Mappa di calore (globale e pre-attentive)

Per ogni osservatore, cioè ogni volta in cui l'esperimento viene ripetuto, l'eye tracker produce in uscita un file di testo contenente tutte le informazioni necessarie per le analisi. I dati sono poi stati ripuliti di dettagli accessori e non utili. Per quanto riguarda l'esperimento eseguito, abbiamo per ogni immagine e per ogni osservatore i seguenti dati:

- un file di testo contenente tutti i punti di osservazione raccolti ogni 15 ms;
- un file excel che è il contenuto del file di testo convertito in formato di tabella per essere agevolmente importato in Matlab.

Con questi dati ci siamo successivamente spostati in ambiente Matlab per eseguire i medesimi passaggi che abbiamo svolto con *MASSVIS* e che sono descritti nel capitolo precedente. In particolare, i risultati grafici mostrati in questa sezione sono:

- nella parte superiore in figura 3.3 l'immagine *graph0* con sovrapposti i punti di fissazione a sinistra e a destra la relativa mappa di calore associata; al di sotto le stesse due figure con solo i punti presi nei primi 2,5 s (per osservare la presenza o meno di qualche differenza nella fase preattentive); in fondo la mappa di salienza generata dall'algoritmo di Matzen;
- in figura 3.4 le stesse informazioni precedenti, ma relative all'immagine *graph5*;
- in figura 3.5 le informazioni relative rispettivamente alle immagini *graph11* e *graph6* in cui non si è analizzata la fase preattentive semplicemente per non caricare di ulteriori immagini il documento (tutto il materiale completo è comunque disponibile ed è descritto nella sezione 4.3).

Le osservazioni e i commenti a queste figure sono lasciati nella prossima sezione in cui vengono analizzati i risultati grafici delle analisi che sono appena state raccontate e anche i dati delle prestazioni con le metriche.

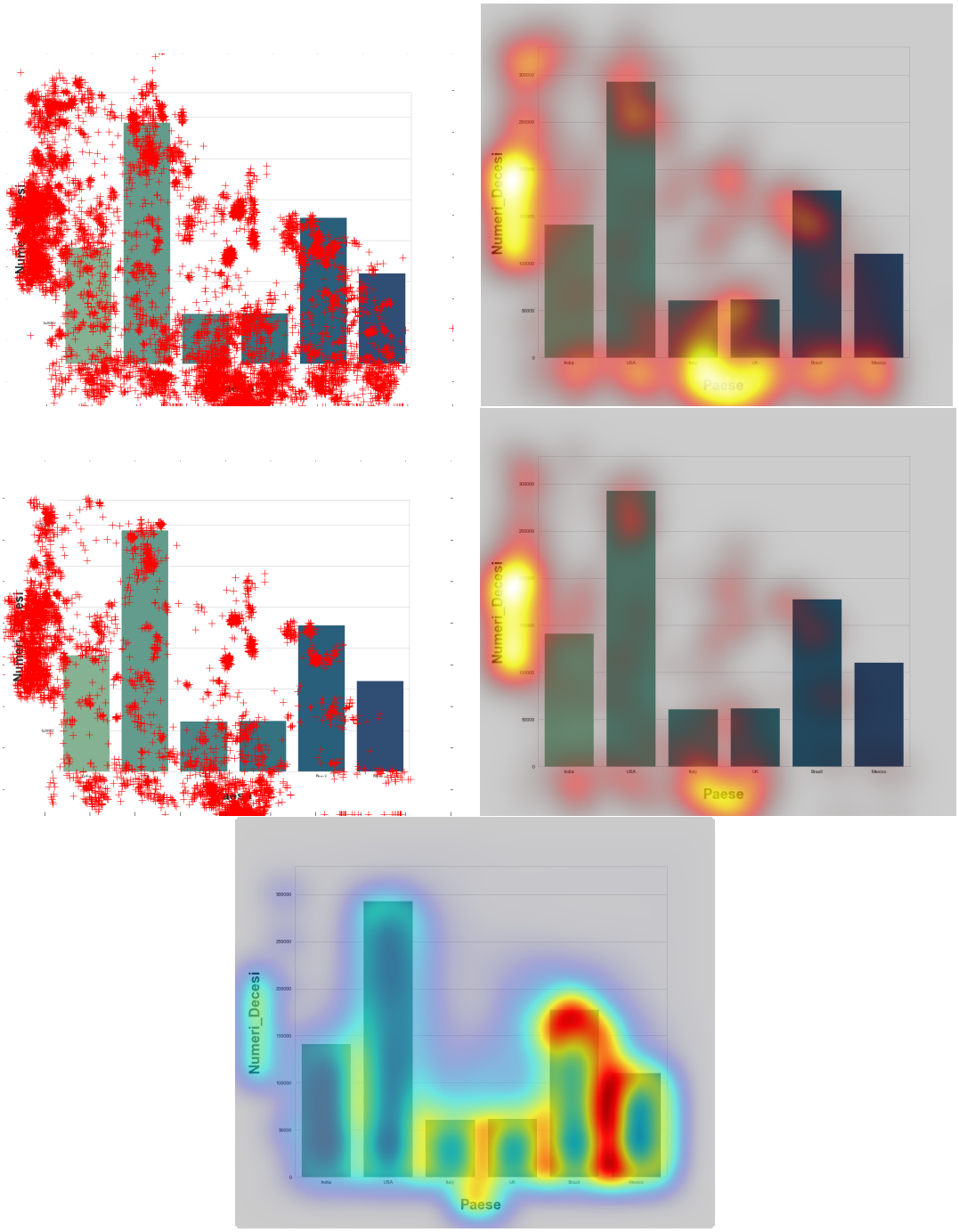

Figura 3.4. Punti di fissazione - Mappa di calore (globale e pre-attentive)

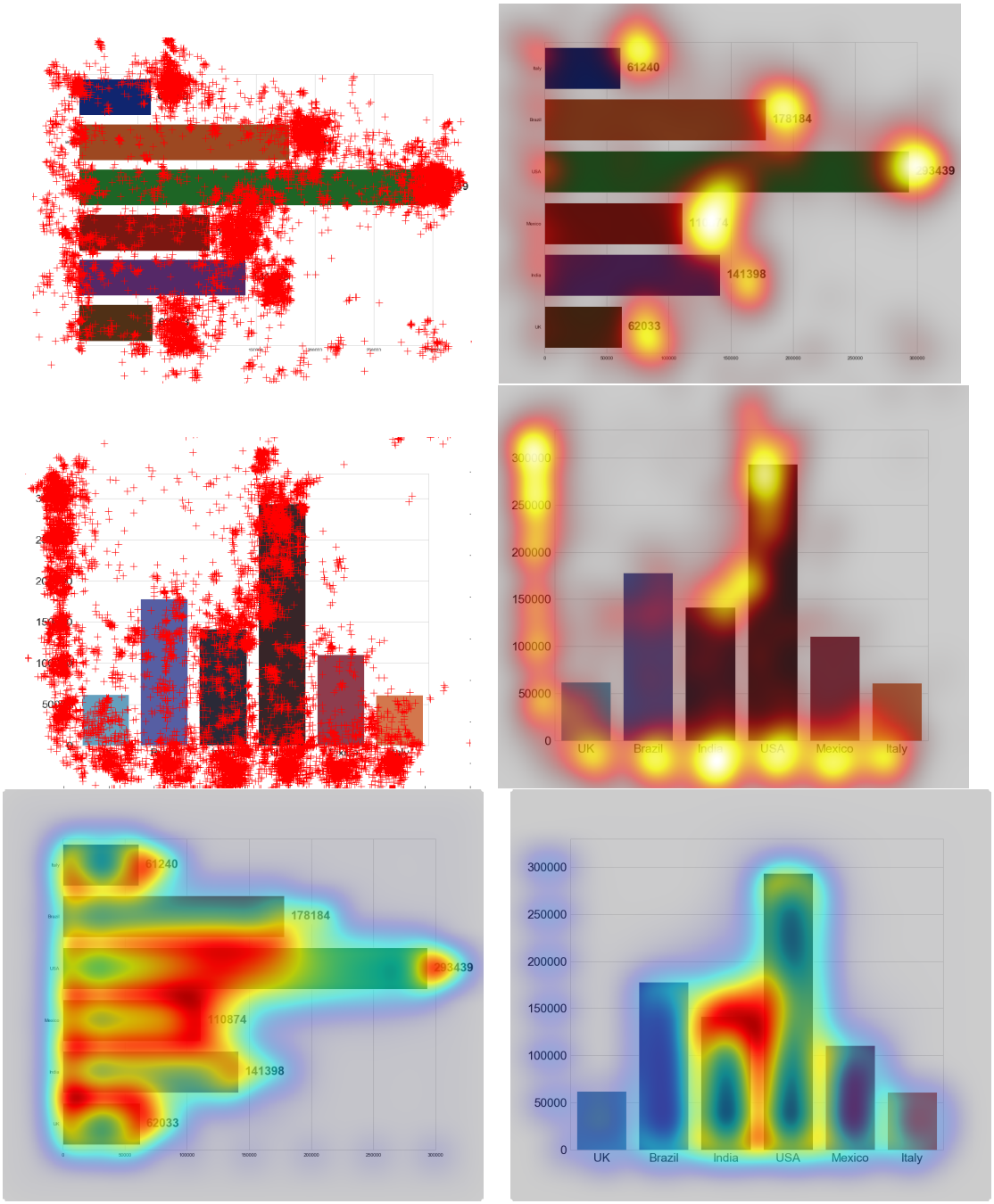

Figura 3.5. Punti di fissazione - Mappa di calore

3.3 Risultati

In questa parte andiamo a commentare sia a livello grafico che a livello numerico i risultati che sono emersi dalle analisi effettuate sui dati raccolti in seguito al nostro esperimento.

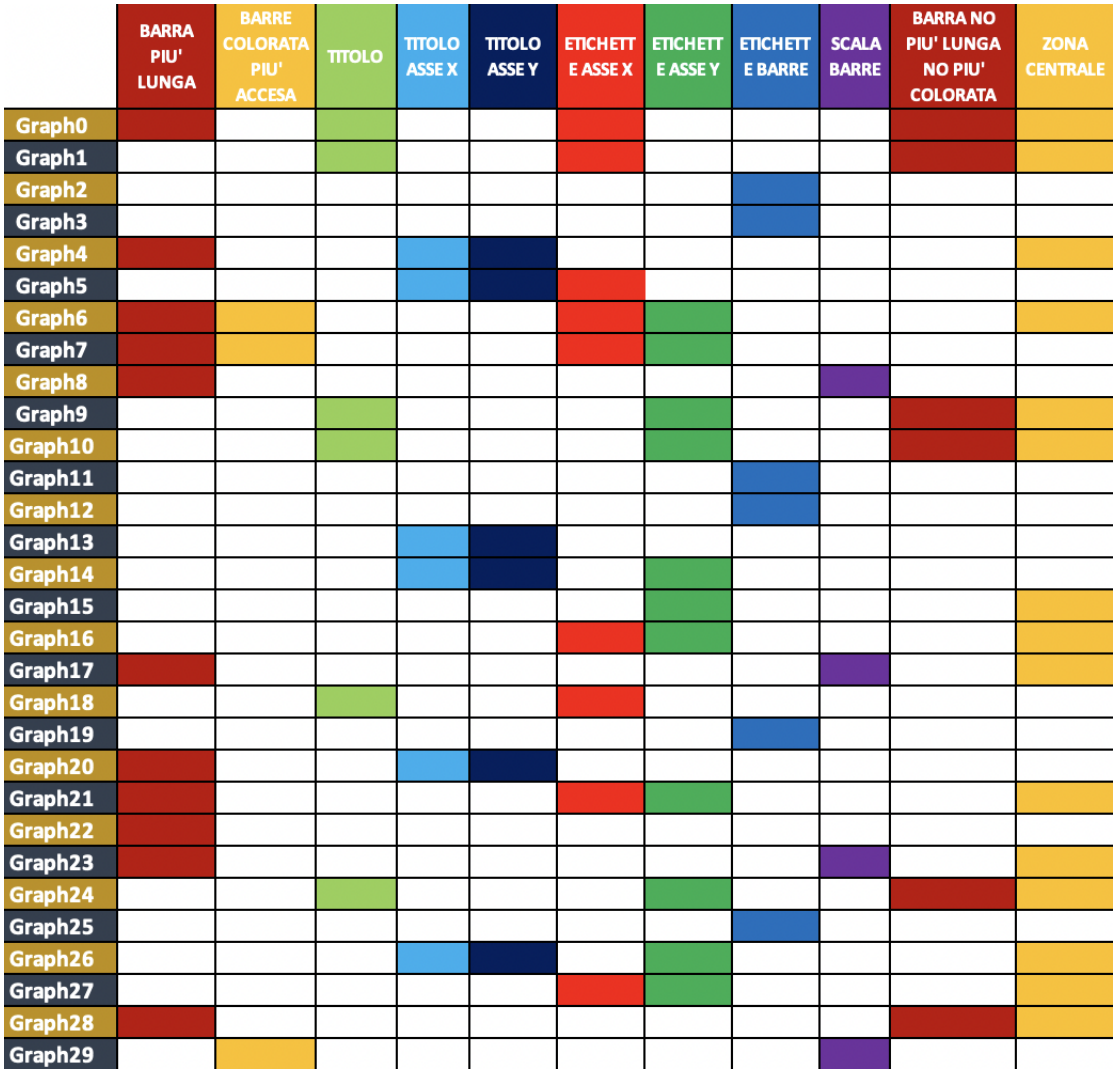

Figura 3.6. Distribuzione mappa di calore esperimento eye-tracker

Dal punto di vista grafico, osservando le mappe di calore, abbiamo evidenziato alcune caratteristiche sulle quali l'attenzione si è prevalentemente concentrata che si ripetono tra le immagini. Sono mostrate in maniera riassuntiva in tabella 3.6 in cui per ognuna delle immagini del nostro dataset sono segnalate le caratteristiche

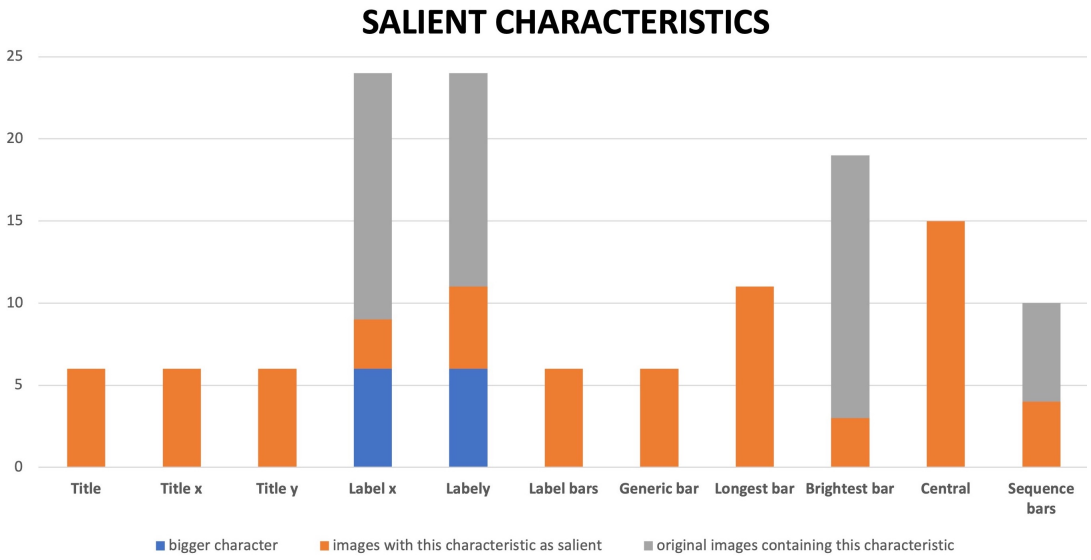

Figura 3.7. Istogramma distribuzione mappa di calore esperimento eye-tracker

di *saliency* ad essa associate in base a quello che gli osservatori hanno ravvisato. Le stesse caratteristiche sono riassunte globalmente nel grafico in figura 3.7 e sono:

- *barra più lunga*: significa che l'attenzione degli osservatori si è concentrata sulla barra più lunga presente nell'immagine. Questo fatto si è verificato in 11 immagini;

- *barra colorata più accesa*: l'attenzione si è concentrata sulla barra con il colore più acceso. 19 sono le immagini originali che avevano una barra colorata in maniera più accesa di altre (colonna grigia 3.7) e in 3 casi (colonna arancione) l'attenzione è ricaduta su di essa;

- *titolo*: l'attenzione è concentrata sul titolo in alto nell'immagine (se presente). In riferimento alla tabella in figura 3.2 possiamo osservare come in 6 immagini originali del nostro dataset fosse presente il titolo e in ciascuna di esse l'attenzione è ricaduta su questa caratteristica;

- *titolo asse x*: l'attenzione è concentrata sul titolo dell'asse x. In riferimento alla tabella in figura 3.2 vale lo stesso discorso della caratteristica precedente;

- *titolo asse y*: l'attenzione è concentrata sul titolo dell'asse y. In riferimento alla tabella in figura 3.2 vale lo stesso discorso delle due caratteristiche precedenti;

- *etichette asse x*: significa che l'attenzione si è concentrata sulle etichette (numeriche oppure di testo) presenti sull'asse x. In 24 immagini originali erano presenti le etichette sull'asse x (colonna grigia 3.7), in 6 di queste il carattere era più grande (colonna blu) e in 9 casi (colonna arancione) l'attenzione è ricaduta su questa caratteristica;

- *etichette asse y*: l'attenzione si è concentrata sulle etichette (numeriche oppure di testo) presenti sull'asse y. In 24 immagini originali erano presenti le etichette sull'asse y (colonna grigia 3.7), in 6 di queste il carattere era più grande (colonna blu) e in 11 casi (colonna arancione) l'attenzione è ricaduta su questa caratteristica;

- *etichette barre*: la saliency è concentrata sulle etichette numeriche delle barre. Consultando la tabella in figura 3.2 si può notare come in tutte le immagini in cui erano presenti le etichette sulle barre, parte dell'attenzione si sia focalizzata su di loro;

- *scala barre*: significa che la saliency si è distribuita un pò su tutte le barre. In 10 immagini originali le barre erano disposte in ordine crescente di lunghezza (colonna grigia 3.7) e in 4 casi (colonna arancione) l'attenzione si è distribuita in maniera abbastanza uniforme su tutte;

- *barra no più lunga no più colorata*: significa che parte dell'attenzione si è concentrata sulla barra che non è nè la più lunga presente nell'immagine e neanche quella con il colore più acceso. Questo fatto si è verificato in 6 casi;

- *zona centrale*: l'attenzione si è focalizzata nel centro dell'immagine. Questo fatto si è verificato in 15 casi e potrebbe essere dovuto ad un ritardo dell'osservatore a muoversi con l'occhio per cercare informazioni.

I risultati appena discussi testimoniano ancora una volta il fatto che il testo quando è presente, in particolare quando si tratta del titolo, attira gran parte dell'attenzione su di esso, ma a questa caratteristica si sono aggiunte nuove indicazioni per quanto riguarda anche il colore con cui le informazioni vengono rappresentate oppure la disposizione che potrebbero essere sviluppate in un lavoro futuro.

Concentrandoci invece sul commento dei risultati da un punto di vista numerico, nella tabella 3.2 sono riassunti (nello stesso formato già adottato nel capitolo precedente) i risultati delle metriche a livello di percentuali di vittorie tra i vari algoritmi per quanto riguarda le immagini del nostro dataset utilizzato per l'esperimento. In questa tabella, tutte e 30 le immagini sono state analizzate insieme e ciò che se ne deduce è una prevalenza di Matzen rispetto all'algoritmo di Itti e a quello di Itti modificato, come del resto era già anche stato riscontrato nel capitolo precedente sulle analisi con le immagini di *MASSVIS*. Soffermandosi sulle

	AUC-Borji	AUC-Judd	AUC-S.	CC	EMD	KL	NSS	SIM
ITTI OR	40%	46,67%	56,67%	26,67%	63,33%	63,33%	26,67%	36,67%
ITTI MOD	60%	50%	43,33%	73,33%	36,67%	36,67%	73,33%	63,33%
PAREGGIO	0%	3,33%	0%	0%	0%	0%	0%	0%
	AUC-Borji	AUC-Judd	AUC-S.	CC	EMD	KL	NSS	SIM
MATZEN	46,67%	86,67%	40%	83,33%	10%	20%	83,33%	76,67%
ITTI OR	53,33%	13,33%	60%	16,67%	90%	80%	16,67%	23,33%
PAREGGIO	0%	0%	0%	0%	0%	0%	0%	0%
	AUC-Borji	AUC-Judd	AUC-S.	CC	EMD	KL	NSS	SIM
MATZEN	46,67%	76,67%	60%	73,33%	6,67%	16,67%	76,67%	66,67%
ITTI MOD	53,33%	10%	40%	13,34%	80%	70%	10%	20%
PAREGGIO	0%	13,33%	0%	13,33%	13,33%	13,33%	13,33%	13,33%

Tabella 3.2. Risultati metriche globali esperimento eye-tracker

percentuali, possiamo osservare come non in tutte le metriche però Matzen risulti effettivamente più efficiente: ad esempio rispetto alla metrica *AUC-Borji* ha delle prestazioni inferiori sia rispetto ad Itti originale che rispetto ad Itti modificato; in più confrontandolo ad Itti originale ha prestazioni inferiori anche in relazione alla metrica *AUC-Shuffled*. Ricordiamo che le metriche *AUC* misurano di fatto la stessa cosa, ciò che cambia è la gestione delle curve *ROC* e dei falsi positivi. Quindi, dato che rispetto ad *AUC-Judd* Matzen rimane il migliore in entrambi i casi, la spiegazione può essere che la gestione degli errori nelle altre due metriche sia più penalizzante. In qualunque caso, confrontando queste percentuali con quelle in tabella 2.1 del capitolo precedente, notiamo come le prestazioni di Matzen siano diminuite sensibilmente. Questo calo è motivato dal fatto che la natura delle immagini, anche se sempre di grafici di dati si tratta sia in *MASSVIS* che nel nostro dataset, influisce sulle performance.

Alla luce delle analisi effettuate nella sezione 3.2 di questo documento, ci possiamo rendere conto di come appena nelle immagini non sia presente più il titolo principale in alto, Matzen abbia delle difficoltà. Questo è testimoniato dalle figure 3.4 e 3.5, in cui si vede che le mappe di salienza generate da Matzen non sono coerenti con le zone effettivamente salienti generate dai dati raccolti dalle persone che hanno partecipato all'esperimento. Per questo motivo sono presentate qui di seguito altre due tabelle che mostrano i risultati delle metriche solo per le immagini senza titolo (tabella 3.3) oppure solo per le immagini con titolo (tabella 3.4). Si può osservare come le percentuali di vittoria di Matzen cambino rispetto alla situazione globale, ma come comunque globalmente rimanga l'algoritmo più efficiente tra i tre analizzati e discussi in questo documento. In particolare possiamo affermare che, la sconfitta di globale di Matzen nella metrica *AUC-Borji* rispetto

	AUC-Borji	AUC-Judd	AUC-S.	CC	EMD	KL	NSS	SIM
ITTI OR	45,83%	58,33%	62,5%	33,33%	70,83%	54,17%	33,33%	33,33%
ITTI MOD	54,17%	37,5%	37,5%	66,67%	29,17%	45,83%	66,67%	66,67%
PAREGGIO	0%	4,17%	0%	0%	0%	0%	0%	0%
	AUC-Borji	AUC-Judd	AUC-S.	CC	EMD	KL	NSS	SIM
MATZEN	54,17%	83,33%	50%	79,17%	8,33%	25%	79,17%	79,17%
ITTI OR	45,83%	16,67%	50%	20,83%	91,67%	75%	20,83%	20,83%
PAREGGIO	0%	0%	0%	0%	0%	0%	0%	0%
	AUC-Borji	AUC-Judd	AUC-S.	CC	EMD	KL	NSS	SIM
MATZEN	54,17%	70,83%	66,67%	66,67%	8,33%	16,67%	70,83%	66,67%
ITTI MOD	45,83%	12,5%	33,33%	16,66%	75%	66,66%	12,5%	16,66%
PAREGGIO	0%	16,67%	0%	16,67%	16,67%	16,67%	16,67%	16,67%

Tabella 3.3. Risultati metriche immagini senza titolo esperimento eye-tracker

ad entrambi gli altri due modelli e la sconfitta nella metrica *AUC-Shuffled* solo rispetto ad Itti originale, sono causate in misura maggiore dalle prestazioni sulle immagini con titolo. Questo conferma il fatto che ciò che può inizialmente apparire a livello grafico, ovvero che Matzen per le immagini con titolo centri correttamente le regioni più salienti, non è detto che si rispecchi tale e quale anche a livello numerico.

	AUC-Borji	AUC-Judd	AUC-S.	CC	EMD	KL	NSS	SIM
ITTI OR	16,67%	0%	33,33%	0%	33,33%	100%	0%	50%
ITTI MOD	83,33%	100%	66,67%	100%	66,67%	0%	100%	50%
PAREGGIO	0%	0%	0%	0%	0%	0%	0%	0%
	AUC-Borji	AUC-Judd	AUC-S.	CC	EMD	KL	NSS	SIM
MATZEN	16,67%	100%	0%	100%	16,67%	0%	100%	66,67%
ITTI OR	83,33%	0%	100%	0%	83,33%	100%	0%	33,33%
PAREGGIO	0%	0%	0%	0%	0%	0%	0%	0%
	AUC-Borji	AUC-Judd	AUC-S.	CC	EMD	KL	NSS	SIM
MATZEN	16,67%	100%	33,33%	100%	0%	16,67%	100%	66,67%
ITTI MOD	83,33%	0%	66,67%	0%	100%	83,33%	0%	33,33%
PAREGGIO	0%	0%	0%	0%	0%	0%	0%	0%

Tabella 3.4. Risultati metriche immagini con titolo esperimento eye-tracker

L'ultima situazione che rimane da commentare è quella legata alla fase che abbiamo chiamato *preattentive* (primi 2,5 secondi di osservazione) sulle immagini del nostro dataset. Facciamo dunque riferimento alle immagini centrali in figura 3.3 e in figura 3.4. Rispetto al caso in cui sono conservati tutti i punti dell'osservazione

(immagini in alto), possiamo notare come ci si concentri ulteriormente sui titoli rispetto alle altre parti grafiche presenti sull'immagine. Ciò significa che ad un primo impatto visivo, lo sguardo di chi ha osservato è ricaduto su queste parti di testo e solo in seguito si è concentrato anche sull'analizzare le informazioni contenute nelle barre.

Capitolo 4

Conclusioni

4.1 Cosa abbiamo scoperto

In conclusione a questo lavoro elenchiamo le informazioni che abbiamo scoperto e confermato riguardo al funzionamento degli algoritmi predittivi della *saliency* che abbiamo analizzato.

Sicuramente abbiamo verificato che in linea generale Matzen performa meglio dell'algoritmo di Itti, sia per quanto riguarda il dataset di *MASSVIS* e sia sulle immagini utilizzate nel nostro esperimento nonostante un calo di prestazioni.

L'argomento sul quale si potrebbe aprire un dibattito è il *testo* e in particolare su come questo elemento venga riconosciuto sull'immagine. Nella sezione successiva, verranno presentati alcuni possibili accorgimenti che potranno essere presi per cercare di ovviare ad alcune mancanze che sono state rivelate in relazione a questo aspetto. In figura 4.1 sono mostrati i risultati solo dell'analisi del testo effettuata dal modello di Matzen per le cinque immagini del dataset utilizzato nel nostro esperimento che erano presenti in 3.1. Possiamo osservare che:

- per la figura che non contiene nessuna parte di testo, correttamente non viene rilevato nulla (centro a destra);
- le porzioni di testo che sono in carattere più grande vengono correttamente individuate (il titolo per la figura in alto a sinistra, le etichette sulle barre nella figura in alto a destra, le etichette sugli assi della figura in centro a sinistra e infine i titoli sugli assi in quella in basso);
- le porzioni di testo leggermente in carattere più piccolo non vengono quasi riconosciute (ad esempio nell'immagine con il titolo, le etichette sugli assi si intravedono appena in ciò che rimane dopo l'analisi).

Nell'immagine in basso a sinistra in 2.2 ricordiamo che è presente un esempio di come il testo venga individuato sul dataset di *MASSVIS*. Confrontando questa

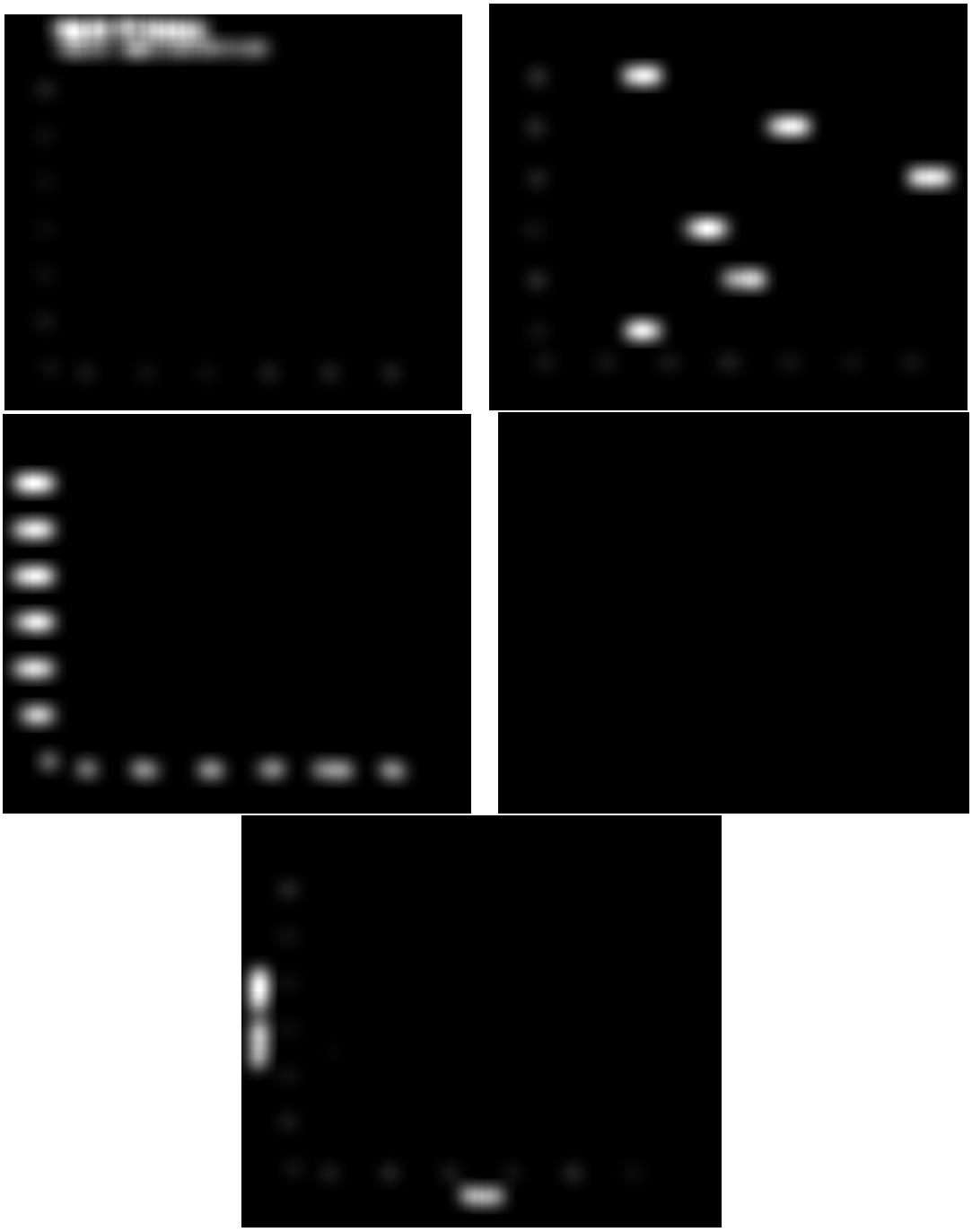

Figura 4.1. Analisi testo esperimento eye-tracker

con l'immagine originale (in alto a sinistra in 2.3), possiamo accorgerci di come tutto il testo venga praticamente riconosciuto e rappresentato correttamente. La giustificazione possibile a questo fatto è che rispetto alle nostre immagini, la maggior parte di quelle che sono contenute in *MASSVIS* hanno il carattere del testo che non subisce variazioni sensibili ed è abbastanza uniforme su tutta l'immagine. Questo rende maggiormente possibile che qualche parte non venga riconosciuta.

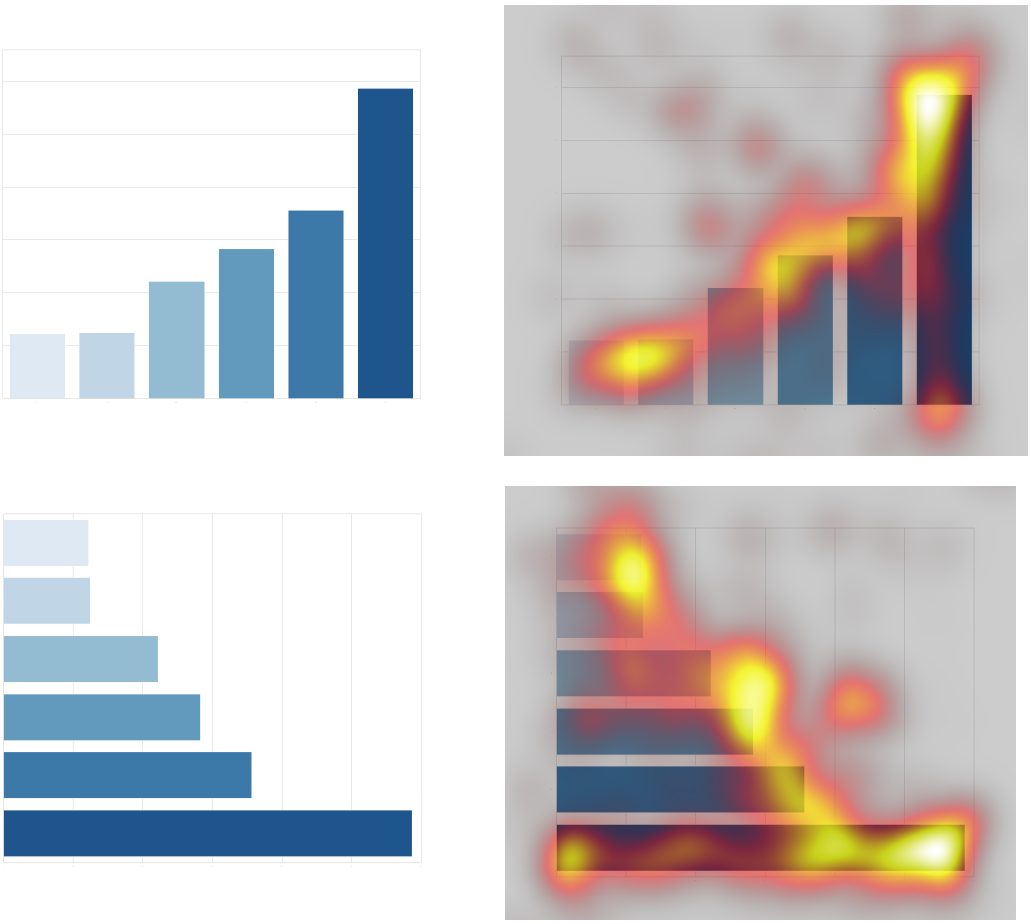

Figura 4.2. Confronto verticale orizzontale mappa di calore

Nella sezione conclusiva di questo documento è descritto tutto il materiale che abbiamo elaborato e nella sezione *EXPERIMENT*, relativa all'esperimento che abbiamo svolto tramite l'eye tracker, è presente una sottocartella *DIREZIONE* che merita una discussione. Questa nasce dall'analisi sulle mappe di calore generate e sovrapposte alle immagini originali. Scorrendo questi grafici ci è sembrato a primo impatto che la caratteristica relativa all'orientamento delle barre potesse influire sull'attenzione dell'occhio, anche nelle immagini che contengono le stesse identiche

informazioni e per le quali varia proprio solo questo aspetto. Ad esempio in ognuna

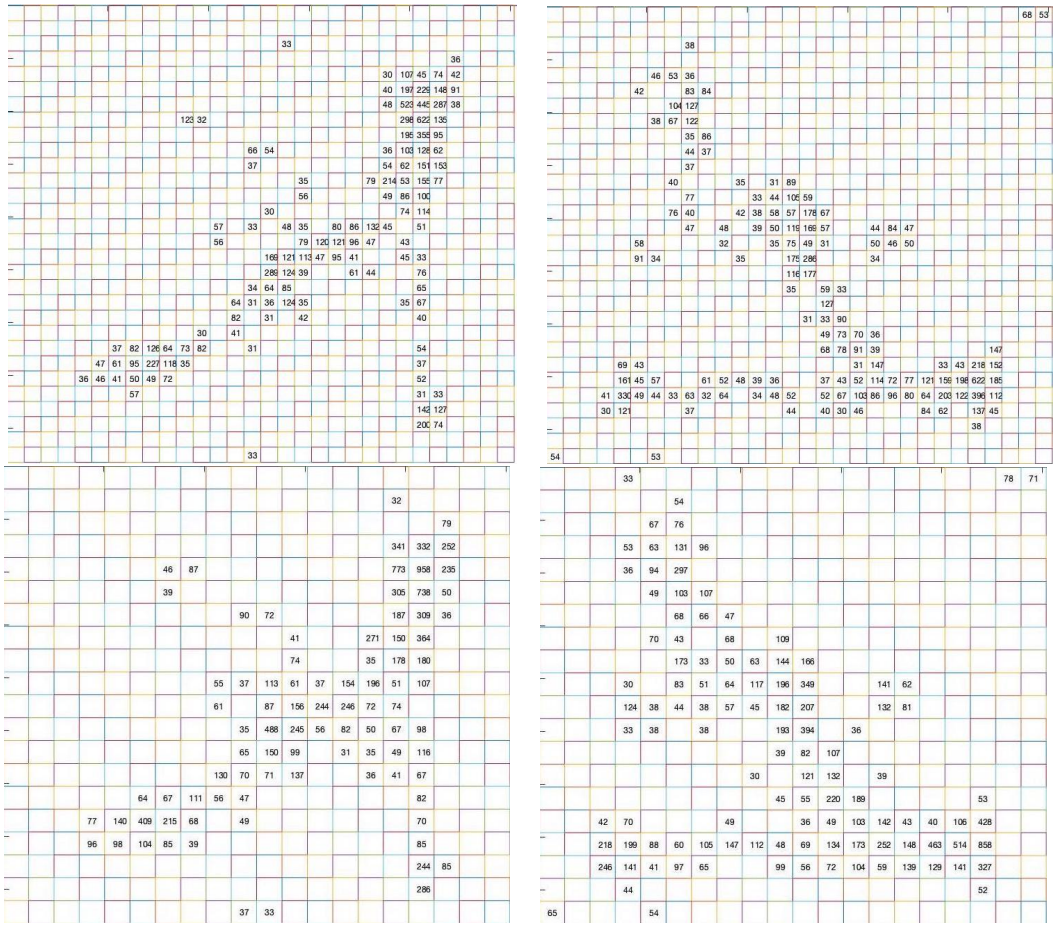

Figura 4.3. Confronto verticale orizzontale distribuzione

delle figure 4.2 e 4.4 abbiamo a sinistra una coppia di immagini che presentano le stesse identiche proprietà grafiche (di quelle elencate nella sezione 3.1) ad eccezione dell'orientazione delle barre: in una sono verticali e nell'altra orizzontali. Osservando solo le mappe di calore sovrapposte alle immagini originali (presenti nella parte destra) potrebbe sembrare che dove le barre sono disposte orrizzontalemente ci sia una maggiore confusione da parte dell'occhio degli osservatori. Per poter confermare oppure smentire questa ipotesi ci è venuto in mente di suddividere l'immagine in rettangoli andando a calcolare quanti punti di osservazione finissero all'interno di ognuno di essi per valutare globalmente la dispersione sull'immagine. Sono state effettuate due tipi di suddivisioni: una con 900 e una con 400 rettangoli. Successivamente è stato riportato nel centro di ciascun rettangolo il numero di punti di osservazione contenuti al suo interno secondo alcune soglie

(al di sotto delle quali nel rettangolo non è stato scritto nulla): per il frazionamento in 900 rettangoli sono state utilizzate le soglie 10, 20 e 30, mentre con 400 rettangoli le soglie usate sono state 30 e 40. I valori delle soglie sono stati scelti considerando il numero medio di punti che ogni rettangolo avrebbe dovuto avere partendo dal fatto che per il nostro esperimento abbiamo avuto una raccolta di circa 15000 punti per ogni immagine (mettendo insieme tutti e 62 gli osservatori). In figura 4.3 sono mostrati i risultati di queste analisi per le due immagini in 4.2:

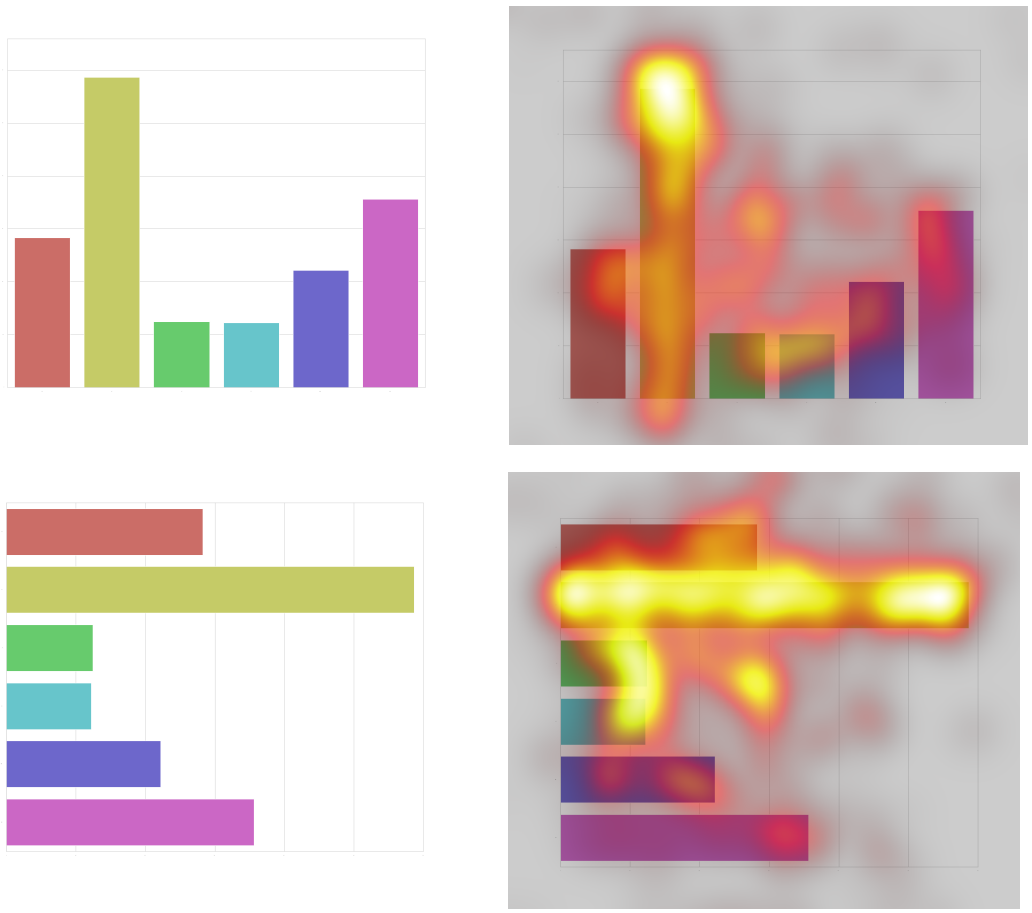

Figura 4.4. Confronto verticale orizzontale mappa di calore

in alto il confronto nella suddivisione in 900 rettangoli con soglia 30 e in basso nella suddivisione in 400 rettangoli sempre con soglia 30. Purtroppo l'ipotesi non si può certo dire confermata perchè non si nota una particolare differenza nella dispersione dei punti tra orizzontale e verticale. Neanche per le altre soglie o per le altre immagini (controllare nel materiale allegato) si riscontra qualche anomalia significativa.

Nella sezione conclusiva viene elencata anche la presenza di una cartella chiamata *PREATTENTIVE* in cui sono state eseguite le stesse procedure appena descritte, ma concentrandosi solo nel primo secondo di osservazione. Anche in questo caso non è stata riscontrata nessuna particolare differenza nella dispersione dei punti tra le immagini con le barre distribuite orizontalmente e quelle con le barre distribuite verticalmente.

4.2 Obiettivi futuri

In questa penultima parte andiamo a presentare quelle che sono delle proposte, degli spunti, su come si potrebbe agire in futuro se qualcuno volesse prendere questo lavoro e provare a continuarlo.

L'idea di effettuare un esperimento nostro in cui potevamo creare e controllare noi le immagini proposte era ottima, ma guardando oltre ecco come si potrebbe agire meglio:

- prelevare un campione di osservatori più vario e meglio distribuito dato che questo è stato probabilmente un difetto che ha accompagnato i nostri risultati. Come già raccontato in precedenza, ci siamo dovuti un pò accontentare perchè nel corso delle nostre indagini purtroppo l'emergenza sanitaria non ci ha permesso di effettuare l'esperimento nelle condizioni in cui avevamo inizialmente pensato;

- effettuare un esperimento ad hoc per ognuna delle caratteristiche presentate nella sezione 3.1. Il nostro esperimento è stato legato a 30 immagini sulle quali sono state distribuite circa otto caratteristiche principali. Sarebbe interessante concentrarsi su immagini che si differenzino tra di loro solo per una o massimo due caratteristiche per poter apprezzare meglio eventuali cambiamenti nelle mappe di calore. Ad esempio in merito a quanto discusso nella sezione 4.1 si potrebbero creare delle coppie di figure in cui cambia solo la distribuzione tra verticale ed orizzontale senza far intervenire ulteriori variazioni (come i colori o cose di questo tipo), ma piuttosto cambiando solo l'ordine delle barre per osservare se influisce su una differente dispersione dei punti di osservazione.

Un altro punto sul quale ci si potrebbe concentrare è andare a mettere mano sull'algoritmo di Matzen ed è sicuramente un compito non facile. Alla luce di quanto è emerso dai risultati grafici raccontati in questo documento, in particolare in merito al nostro esperimento, abbiamo notato come il testo rappresenti il punto principale su cui focalizzarsi dato che le parti grafiche passano in secondo piano quando il testo è presente. Un'errata analisi sul testo porta molto probabilmente ad una generazione della mappa di salienza non corretta. In particolare, dato che

quando il testo ha caratteri di dimensioni differente Matzen (a livello grafico con la sua mappa di salienza generata) fa fatica a riconoscerne delle porzioni, si potrebbe forse migliorare l'algoritmo andando ad agire sulla ricerca nella figura di porzioni di testo che via via sono sempre più piccole. Il focus potrebbe essere lavorare su quella parte dell'algoritmo che nel dettaglio inizia dalla ricerca di quello che può essere il titolo (tipicamente con carattere più grande se è presente) fino a scendere a scritte che possono essere ad esempio delle legende. Come ultima cosa, è importante però che con le eventuali modifiche l'algoritmo non faccia come succede in alto a destra della figura 2.7, dove quando gran parte del testo è della stessa dimensione, allora la mappa di calore si distribuisce uniformemente su di essa. Quando ci si trova in questa situazione (distese porzioni di testo della stessa dimensione) probabilmente bisognerebbe andare valutare la disposizione in base a dati statistici e alla durata dell'esperimento, cioè per quanti secondi l'osservatore guarda l'immagine. Ad esempio, se ci si dovesse trovare sia con del testo in alto a sinistra e sia in alto a destra sull'immagine, allora con buona probabilità il testo a sinistra sarà più rilevante se i caratteri sono della stessa dimensione (vista la maniera classica di lettura del testo da sinistra verso destra). L'importante è evitare di evidenziare tutto saliente allo stesso modo, altrimenti l'obiettivo dell'algoritmo perde un pò di significato.

4.3 Materiale allegato

Durante questi mesi di lavoro alla tesi, è stato prodotto molto materiale. Quando le analisi si sono concluse è stato riorganizzato in una una struttura ad albero (alla quale si può accedere tramite questo link) che è così composta:

- *EXPERIMENT*: questa cartella contiene tutti i riferimenti per quanto riguarda l'esperimento con eye tracker che abbiamo effettuato sulle 30 immagini create da noi. Al suo interno ha tre sottocartelle:

    1. *DIREZIONE*: contiene le informazioni per quanto riguarda il controllo della distribuzione dei punti nelle immagini verticali e orizzontali. E' suddivisa in due sottocartelle:

        (a) *DIREZIONE_GLOBALE*: contiene le cartelle per il controllo di ogni singola coppia di immagini. Ad esempio la *0_9* contiene il confronto diretto tra l'immagine *0* e quella a lei associata che è la numero *9*. In ognuna di queste è contenuta una cartella per ognuna delle due immagini della coppia. Ad esempio in *0_9* è contenuta sia la cartella *0* che la cartella *9*. Ognuna di queste contiene le seguenti informazioni:

i. *FILE_TEXT_400_graph*: l'immagine è stata suddivisa in 400 rettangoli uguali numerati orizzontalmente e in questo file viene riportato in ogni rettangolo quanti punti di fissazione sono contenuti all'interno di essi;
ii. *FILE_TEXT_900_graph*: l'immagine è stata suddivisa in 900 rettangoli uguali numerati orizzontalmente e in questo file viene riportato in ogni rettangolo quanti punti di fissazione sono contenuti all'interno di essi;
iii. *graph*: l'immagine originale;
iv. *IMAGE_HEAT_MAP_62_graph*: questa è l'immagine originale usata nell'esperimento alla quale è stata sovrapposta la mappa di calore mettendo insieme i dati di tutti gli osservatori;
v. *RETTANGOLO_400_30_graph*: immagine contenente solo i 400 rettangoli in cui l'immagine è stata suddivisa con all'interno scritto il numero di punti di fissazione (solo per quelli che ne hanno almeno 30);
vi. *RETTANGOLO_400_40_graph*: immagine contenente solo i 400 rettangoli in cui l'immagine è stata suddivisa con all'interno scritto il numero di punti di fissazione (solo per quelli che ne hanno almeno 40);
vii. *RETTANGOLO_900_10_graph*: immagine contenente solo i 900 rettangoli in cui l'immagine è stata suddivisa con all'interno scritto il numero di punti di fissazione (solo per quelli che ne hanno almeno 10);
viii. *RETTANGOLO_900_20_graph*: immagine contenente solo i 900 rettangoli in cui l'immagine è stata suddivisa con all'interno scritto il numero di punti di fissazione (solo per quelli che ne hanno almeno 20);
ix. *RETTANGOLO_900_30_graph*: immagine contenente solo i 900 rettangoli in cui l'immagine è stata suddivisa con all'interno scritto il numero di punti di fissazione (solo per quelli che ne hanno almeno 30).

(b) *DIREZIONE_PREATTENTIVE*: contiene le stesse informazioni di quella precedente, ma è concentrata solo nel primo secondo di fissazione. In pratica per ogni osservatore abbiamo conservato solo i punti relativi al primo secondo di fissazione. Nelle singole cartelle di ogni immagine, non c'è l'informazione sulla mappa di calore, ma è inserita in più l'immagine *RETTANGOLO_900_graph0* che contiene l'informazione sulla distribuzione dei punti di fissazione senza tenere conto di nessuna soglia con cui scrivere i numeri nei rettangoli.

2. *EYETRACKER*: contiene una cartella per ognuna delle immagini del nostro dataset con i risultati dell'esperimento con l'eye tracker. Nella cartella di ogni immagine sono contenute le seguenti informazioni:
   (a) *Combination_graph*: è la combinazione tra quello che genera Itti modificato e l'analisi del testo sull'immagine in questione. Da questa viene poi generata la mappa di calore di Matzen;
   (b) *FIRST_IMAGE_FIXATIONS_62_graph*: è l'immagine originale con sovrapposti i punti di fissazione solo della fase pre-attentive, quindi solo i primi 2,5 s;
   (c) *FIRST_IMAGE_HEAT_MAP_62_graph*: è l'immagine originale con sovrapposta la mappa di calore solo della fase pre-attentive;
   (d) *graph*: è l'immagine originale;
   (e) *IMAGE_FIXATIONS_62_graph*: è l'immagine originale con sovrapposti tutti i punti di fissazione;
   (f) *IMAGE_HEAT_MAP_62_graph*: è l'immagine originale con sovrapposta la mappa di calore generata da tutti i punti di fissazione;
   (g) *ITTI_graph*: è quello che l'algoritmo di Itti modificato genera;
   (h) *ITTI_OR_graph*: è quello che l'algoritmo di Itti originale genera;
   (i) *mySalience_graph*: è l'immagine originale con sovrapposta la mappa di calore generata da Matzen;
   (j) una cartella *oss* per ognuno dei 62 osservatori che sono stati sottoposti all'esperimento: in ognuna di queste cartelle è contenuto il file di testo *subject...* che contiene tuti i punti di fissazione per quel determinato osservatore su quell'immagine, mentre il file excel *Cartel1* è la conversione in formato di tabella del file di testo precedente in maniera tale che sia agile importarlo in Matlab;
   (k) *TextSaliency_graph*: è quello che viene generato solo dall'analisi del testo.
3. *ORIGINALI*: contiene tutte e 30 le immagini usate per l'esperimento così come sono state create originariamente, da *graph0* a *graph29*.

- *MASSVIS*: questa cartella contiene tutte le informazioni riguardanti il dataset di *MASSVIS*. Al suo interno contiene quattro sottocartelle:

1. *ANALISI_SEMANTICA*: contiene tutte le informazioni per quanto riguarda l'analisi semantica, quindi poligoni sovrapposti all'immagine, effettuata su ognuna delle 110 immagini di *MASSVIS*. All'interno di questa cartella abbiamo una sottocartella per ognuna delle immagini e all'interno troviamo le seguenti informazioni:

   (a) *immagine*: file di testo fornito da MASSVIS che contiene le delimitazioni dei poligoni i quali rappresentano delle specifiche porzioni sull'immagine (titolo, parte grafica ecc.);
   (b) *immagine.txt*: file di testo prodotto da noi in cui per ogni poligono abbiamo inserito il numero di punti di fissazione che cadono al suo interno;
   (c) *EyeTracking_*: immagine che contiene l'immagine originale alla quale viene sovrapposta la mappa di calore in base ai punti di fissazione;
   (d) *POLY_AND_POINTS_*: Immagine con solo i poligoni disegnati su sfondo bianco e con i punti di fissazione che sono stati rappresentati all'interno del poligono opportuno;
   (e) *POLY_ON_IMG_*: immagine con i poligoni sovrapposti all'immagine.

2. EYETRACKING: contiene le analisi che abbiamo effettuato su *MASSVIS* relative alla mappa di calore vera generata dai punti di fissazione e relative a ciò che l'algoritmo di Itti, Itti modificato e Matzen producono graficamente. In particolare abbiamo:
   (a) *Combination_*: è la combinazione tra quello che genera Itti modificato e l'analisi del testo sull'immagine in questione. Da questa viene poi generata la mappa di calore di Matzen;
   (b) eventuali cartelle *CROP* relative ai tagli compiuti sull'immagine: se presenti, all'interno di queste cartelle si trovano le stesse identiche informazioni di questo elenco letterale a cui si aggiungono *CROP_* che rappresenta l'immagine originale tagliata e invece *OR_* è l'immagine originale;
   (c) *EyeTracking_*: è l'immagine originale con sovrapposta la mappa di calore generata da tutti i punti di fissazione raccolti dagli osservatori;
   (d) *Final_*: è l'immagine originale con sovrapposta la mappa di calore generata da Matzen;
   (e) *ITTI_*: è quello che l'algoritmo di Itti modificato genera;
   (f) *ITTI_OR_*: è quello che l'algoritmo di Itti originale genera;
   (g) *TextSaliency_*: è quello che viene generato solo dall'analisi del testo.
3. *ORIGINALI*: contiene le 110 immagini originali che abbiamo conservato dal dataset di *MASSVIS*.
4. PREATTENTIVE: questa cartella fa riferimento alla mappa di calore dei punti di fissazione della cartella precedente *EYETRACKING*, dove però sono stati utilizzati solo i punti presenti nella fase pre-atttentive. In particolare abbiamo:

(a) *EyeTracking_*: è l'immagine originale con sovrapposta la mappa di calore generata da tutti i punti di fissazione raccolti dagli osservatori;

(b) *EyeTracking_MOD250_*: è l'immagine originale con sovrapposta la mappa di calore generata dai punti di fissazione raccolti dagli osservatori solo nei primi 250 ms;

(c) *EyeTracking_MOD500_*: è l'immagine originale con sovrapposta la mappa di calore generata dai punti di fissazione raccolti dagli osservatori solo nei primi 500 ms;

(d) *EyeTracking_MODNOFIRST250_*: è l'immagine originale con sovrapposta la mappa di calore generata dai punti di fissazione raccolti dagli osservatori escludendo i primi 250 ms e prendendo i 250 ms successivi a quelli esclusi.

- *METRICHE*: questo file excel contiene le informazioni riguardo ai risultati sul calcolo delle metriche sulle varie mappe di saliency prodotte dagli algoritmi con cui abbiamo lavorato (Itti, Itti modificato e Matzen). Nella prima parte vengono analizzate le 110 immagini del dataset di MASSVIS che abbiamo conservato e vengono inseriti i valori delle 8 metriche per ognuno dei 3 algoritmi. Di fianco viene calcolata la differenza tra gli algoritmi. Nella parte sottostante al calcolo delle differenze ci sono delle tabelle riassuntive che permettono di avere risultati più compatti su quale algoritmo performi meglio degli altri in termini di percentuale di vittorie. Al di sotto di questa prima parte abbiamo i risultati dell'analisi sul testo, in cui 11 immagini di MASSVIS sono state tagliate e private di alcune porzioni di testo (per alcune immagini sono stati effettuati anche più tagli). Dopo aver eseguito i tagli, sono stati calcolati nuovamente i valori delle 8 metriche sui 3 algoritmi come in precedenza con sempre delle tabelle riassuntive per valutare le performance a confronto. Dopodiché sono contenuti i risultati dell'esperimento sulle nostre immagini. Per ognuna delle 30 immagini sono stati calcolati i valori delle 8 metriche, calcolate le differenze e messe a confronto in tabelle per confrontare le prestazioni. Al fondo sono presenti due tabelle: la prima descrive la distribuzione delle caratteristiche tra le nostre immagini usate nell'esperimento; la seconda tabella riassume la distribuzione delle caratteristiche nelle mappe di saliency dopo aver analizzato tutti i dati raccolti sugli osservatori.

- *CODICI MATLAB*: in questa cartella sono contenuti sia i codici Matlab relativi alle analisi sul nostro esperimento e sia quelli per le analisi compiute su *MASSVIS*.

Bibliografia

[1] C. Healey and J. Enns, "Attention and Visual Memory in Visualization and Computer Graphics," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 7, pp. 1170-1188, July 2012, doi: 10.1109/TVCG.2011.127.

[2] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259, Nov 1998. doi: 10.1109/34. 730558

[3] L. E. Matzen, M. J. Haass, K. M. Divis, Z. Wang, and A. T. Wilson. Data visualization saliency model: A tool for evaluating abstract data visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):563–573, Jan 2018. doi: 10.1109/TVCG.2017.2743939

[4] A. Borji and L. Itti. State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):185–207, 2013. doi: 10.1109/TPAMI.2012.89

[5] Z. Bylinskii, T. Judd, A. Oliva, A. Torralba, and F. Durand. What do different evaluation metrics tell us about saliency models?, 2017.

[6] Z. Bylinskii, N. W. Kim, P. O'Donovan, S. Alsheikh, S. Madan, H. Pfister, F. Durand, B. Russell, and A. Hertzmann. Learning visual importance for graphic designs and data visualizations. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, UIST '17, p. 57–69. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3126594.3126653

[7] C. Fosco, V. Casser, A. K. Bedi, P. O'Donovan, A. Hertzmann, and Z. Bylinskii. Predicting visual importance across graphic design types. UIST '20. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3379337.3415825

[8] M. J. Haass, A. T. Wilson, L. E. Matzen, and K. M. Divis. Modeling human comprehension of data visualizations. In S. Lackey and R. Shumaker, eds., *Virtual, Augmented and Mixed Reality*, pp. 125–134. Springer International Publishing, Cham, 2016.

[9] J. Zhang and S. Sclaroff. Saliency detection: A boolean map approach. In 2013 *IEEE International Conference on Computer Vision*, pp. 153–160, Dec 2013. doi: 10.1109/ICCV.2013.26

[10] E. Vig, M. Dorr, and D. Cox. Large-scale optimization of hierarchical features for saliency prediction in natural images. In 2014 *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2798–2805, 2014. doi: 10.1109/CVPR.2014.358

[11] M. A. Livingston, L. E. Matzen, A. Harrison, A. Lulushi, M. Daniel, M. Dass, D. Brock, and J. W. Decker. A study of perceptual and cognitive models applied to prediction of eye gaze within statistical graphs. In *ACM Symposium on Applied Perception 2020*, SAP '20. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3385955 .3407931

[12] Pele O., Werman M. (2008) A Linear Time Histogram Metric for Improved SIFT Matching. In: Forsyth D., Torr P., Zisserman A. (eds) Computer Vision – ECCV 2008. ECCV 2008. Lecture Notes in Computer Science, vol 5304. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-88690-7_37

[13] Haass M.J., Wilson A.T., Matzen L.E., Divis K.M. (2016) Modeling Human Comprehension of Data Visualizations. In: Lackey S., Shumaker R. (eds) Virtual, Augmented and Mixed Reality. VAMR 2016. Lecture Notes in Computer Science, vol 9740. Springer, Cham. https://doi.org/10.1007/978-3-319-39907-2_12

[14] S. Math^ot, D. Schreij, and J. Theeuwes. OpenSesame: An open-source, graphical experiment builder for the social sciences. Behavior Research Methods, 44(2):314–324, June 2012. doi: 10.3758/s13428-011-0168-7

[15] Ariely D. Seeing Sets: Representation by Statistical Properties. Psychological Science. 2001;12(2):157-162. doi:10.1111/1467-9280.00327

[16] A. Treisman and G. Gelade, "A Feature-Integration Theory of Attention", *Cognitive Psychology*, vol. 12, pp. 97-136, 1980.

[17] A. Treisman, "Preattentive Processing in Vision", *Computer Vision, Graphics and Image Processing*, vol.31, pp. 156-177, 1985.

[18] B. Julész, "Textons, the Elements of Texture Perception, and Their Interactions", *Nature*, vol. 290, pp. 91-97, 1981.

[19] J. Duncan and G.W. Humphreys, "Visual Search and Stimulus Similarity", *Psychological Rev.*, vol. 96, no. 3, pp. 433-458, 1989.

[20] J.M. Wolfe, K.R. Cave, and S.L. Franzel, "Guided Search: An Alternative to the Feature Integration Model for Visual Search", *J. Experimental Psychology: Human Perception and Performance*, vol. 15, no. 3, pp. 419-433, 1989.

[21] L. Huang and H. Pashler, "A boolean Map Theory of Visual Attention", *Psychological Rev.*, vol. 114, no. 3, pp. 599-631, 2007.

[22] L. Huang, A. Treisman, and H. Pashler, "Characterizing the Limits of Human Visual Awareness", *Science*, vol. 317, pp. 823-825, 2007.

[23] J.M. Wolfe, N. Klempen, and K. Dahlen, "Post Attentive Vision", *J. Experimental Psychology: Human Perception and Performance*, vol. 26, no. 2, pp. 693-716, 2000.

[24] H. Greenspan, S. Belongie, R. Goodman, P. Perona, S. Rakshit, and C.H. Anderson, "Overcomplete Steerable Pyramid Filters and Rotation Invariance," *Proc. IEEE Computer Vision and Pattern Recognition*, pp. 222-228, Seattle, Wash., June 1994.

[25] J Atkinson, "The Developing Visual Brain" Oxford University Press, Oxford, UK., 2002.

[26] L Itti and C. Koch, "Computational modelling of visual attention," *Nature Reviews Neuroscience*, vol. 2, pp. 194-203, 2001.

[27] E. Vig, M. Dorr, and D. Cox, "Large-scale optimization of hierarchical features for saliency prediction in natural images," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2798-2805, 2014.

[28] M. D. Fairchild and R. S. Berns, "Image color-appearance specification through extension of CIELAB" Color Research Application, vol. 18, pp. 178-190, 1993.

[29] L. E. Matzen, M. J. Haass, K. M. Divis and M.C. Stites "Patterns of attention: How data visualizations are read," *Augmented Cognition. Enhancing Cognition and Behavior in Complex Human Environments*, D. D. Schmorrow and C. M. Fidopiastis, eds., pp. 176-191. Springer, 2017.

[30] C. M MacLeod, "Half a century of research on the Stroop effect: An integrative review" *Psychological bulletin*, vol. 109, p. 163, 1991.

[31] J. Matas, O. Chum, M. Urban T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image and Vision Computing*, vol. 22, pp. 761-767, 2004.

[32] M. A. Borkin, Z. Bylinskii, N. W. Kim, C. M Bainbridge, C.S Yeh, D. Borkin, H. Pfister, and A. Oliva, "Beyond Memorability: Visualization Recognition and Recall," *IEEE Trans. Visualization and Computer Graphics*, vol. 22, pp.519-528, 2016.