I cronotopi del volto
ISBN 979-12-218-0270-2
DOI 10.53136/97912218027029
pp. 183-201 (ottobre 2022)

RICAMBIARE LO SGUARDO DELLE MACCHINE DIETRO GLI IMPLICITI DELLA *FACE RECOGNITION* ATTRAVERSO LE *TRAINING IMAGES*

Valentina Manchia

Abstract: Facial recognition technologies, which are increasingly part of our experience of daily interaction with artificial intelligence, are often presented, both by developers and by the companies and institutions that adopt them, as pure utilities in a context of simple and effective automation (Floridi 2019, 2020). The vast field of automated facial recognition (AFR) includes all those technologies that apply algorithms to the human face and facial expressions, from face recognition apps to CCTV and police cameras, decoding features and characteristics through a purely artificial vision. What happens, however, if we focus on the peculiar visual AI's way of seeing and, at the same time, classifying the human face? Which are the (social, political, cultural, and semiotic) implications of automated facial recognition, machinic vision with the human face as its object? A number of artistic projects have turned attention to facial recognition technologies by reflecting on the "invisible" images (Paglen 2019) on which computer vision based on convolutional neural network (CNN) feeds in order to function, the so–called *training images*. After some notes on AFR as a form of artificial intelligence, this paper aims to examine some of such artistic projects, specifically Trevor Paglen's *ImageNet Roulette* and *From "Apple" to "Anomaly"* (2019), considering them a possible way to access a greater understanding of facial recognition as well as computer vision as a peculiar phenomenon of the contemporary visual landscape.

Keywords: artificial intelligence, computer vision, automated facial recognition, ImageNet, semiotics, semantics

Sempre più spesso su smartphone e tablet, in app di e–commerce, in telecamere di sorveglianza a circuito chiuso o nel database delle foto riprese durante rilievi segnaletici operano tecnologie biometriche per il riconoscimento facciale.

Questi e altri esempi di *automated facial recognition* (AFR), così come molte delle forme attuali di *artificial intelligence* (AI) con cui ci

confrontiamo abitualmente — dell'autocompletamento nella scrittura delle e–mail ai suggerimenti di acquisto sui vari profili online — assolvono dei compiti per noi in modo automatico e pressoché invisibile. Dell'AFR e di altre forme di AI, in altre parole, vediamo i risultati (lo sblocco di uno smartphone o il percorso giusto per noi, su una mappa), ma non le vediamo in azione.

Il rilievo, apparentemente banale, ha una qualche importanza dal momento che rimarca un cambiamento importante nel modo in cui pensiamo, oggi, all'AI.

Essa, infatti, non è più l'entità, identificabile e nominabile, incarnata, per così dire, in una "thinking machine" (Turing 1950) pensata come *agency* artificiale, strutturata per imitare il funzionamento del pensiero umano. Così è stato dagli albori della cibernetica, dagli anni '50–'60, in cui "la visione originaria dell'IA aveva come obiettivo quello di insegnare a un computer a svolgere una serie di compiti cognitivi" (Manovich 2018), ovvero la messa a punto di un'*agency* macchinica capace di ottenere un risultato attraverso una serie di regole, di un "mechanized thinking" (Shannon 1950).

Assistiamo invece, oggi, a una vera e propria *normalizzazione* dell'AI (Floridi 2020) che passa attraverso la sua invisibilità: come argomenta Manovich (2018), "una volta che il campo dell'IA risolve un problema e la soluzione è implementata nell'ambito dell'industria, questa non viene più considerata come parte dell'ambito originario. Paradossalmente, associamo all'Intelligenza Artificiale solo i problemi che non sono ancora stati risolti."

Floridi (2020), inoltre, registra la crescente tendenza, in un contesto istituzionale come quello delle *smart cities*, a integrare le AI (AFR compresa) al servizio dei cittadini come pure *utilities*, secondo una logica di "Utility–Fication" delle AI viste come pure *mindless agency*, ovvero sistemi programmati per ottenere risultati.

La formula è interessante perché cerca di superare l'etichetta stessa di *intelligenza artificiale* come *agency* "pensante", descrivendola invece come un "divorzio" (Floridi 2019) — una disgiunzione (semioticamente parlando) — tra la capacità di mirare a un obiettivo specifico (= problema da risolvere) e il bisogno di essere intelligente per farlo (altrimenti detto, tra l'acquisizione di competenze in vista in un fare e questo stesso fare).

La visione di Floridi (2020) è interessante perché se da un lato mette a fuoco la necessità di supervisionare e gestire l'AI in quanto *mindless* (andando a monte, alla scrittura degli algoritmi che ne guidano l'azione), dall'altro lato mostra bene, in controluce, che la vulgata di aziende e istituzioni descrive l'AI come un puro risultato, sul quale non occorre invece guardare agli impliciti.

Proprio per questo può essere importante interrogarsi sugli impliciti della *mind* che muove questa *agency* invisibile, per approfondire il funzionamento della *face recognition* da un punto di vista non solo tecnologico e cercare così di ricambiare il suo sguardo.

Cercheremo di farlo attraverso alcune esplorazioni artistiche che aprono un vertiginoso spaccato, non privo di implicazioni culturali, sociali, politiche e, non da ultimo, semiotiche, sulla visione artificiale.

1. Breve storia della visione macchinica. Qualche nota sull'AFR come forma di intelligenza artificiale

Prima però di spostare l'attenzione sulla *face recognition* e sulle sue dinamiche intrinseche, è il caso di tracciarne un rapido profilo.

A farne una breve storia, l'AFR nasce negli anni '90, con la messa a punto dei primi algoritmi per consentire alle macchine la decodifica del volto. L'AFR, tuttavia, sarebbe impensabile senza l'antecedente apporto della *computer vision*, dapprima grazie a Marvin Minsky e alla sua idea di sperimentare, nel laboratorio di Intelligenza artificiale al MIT di Boston, la connessione tra una fotocamera e un computer (sin dal 1966) e successivamente grazie agli studi e alle ricerche di David Marr, a partire dagli anni '70.

Come l'etichetta di *artificial intelligence* porta molto spesso a dimenticare che le macchine non sono dotate di un'intelligenza del tutto analoga a quella umana, anche quella di *computer vision* nasconde il fatto che nel computer non c'è una vera e propria visione, paragonabile a una visione organica: piuttosto, si tratta un filtraggio di elementi visivi di input, secondo dei parametri prestabiliti, per ottenere, come output, un'elaborazione degli input di partenza.

Già Marr (1982, ma pubblicato postumo), proponendo una via per l'implementazione della *computer vision* a partire dal funzionamento

della vista umana, descrive la visione come un processo che fa corrispondere a una rappresentazione bidimensionale un'interpretazione tridimensionale, e di conseguenza mostra come sia possibile "insegnare" a una macchina, mediante l'applicazione di modelli "mentali" che di fatto traducono tratti 2D in immagini 3D, a interpretare configurazioni di tratti visivi come immagini di oggetti del mondo.

Su questa scia, analogamente, i sistemi di AFR sono in grado di interpretare determinate configurazioni visive di pixel come volti umani, e andando ancora più oltre di analizzare i parametri visivi di un volto secondo determinate caratteristiche, in modo da associare un'immagine a uno specifico referente.

Ora, *image recognition* e AFR sono campi sterminati. Non potremmo nemmeno accennare a tutte le specifiche modalità di *computer vision*, né ai diversi algoritmi per l'AFR messi a punto nel corso degli anni.

In linea generale, possiamo però dire che i sistemi di *face recognition* operano secondo due modi distinti, e si raggruppano di conseguenza in due grandi categorie: i sistemi di *face verification* (o *authentication*) e i sistemi di *face identification* (o *recognition* propriamente detta). Come riassumono Mane e Shah (2019, p. 276), nella loro utile cronistoria sulle tecnologie di *facial analysis*:

> A face recognition system automatically identifies the face present in an image or video. It can operate in one of the following two modes: face verification (or authentication), and face identification (or recognition). For verification, we have an input image and we need to check if this image matches one specific in a given database (one–to–one match). On the other hand, with identification we have to compare a given image with all the images in the database (one–to–many match).

Queste due modalità di riconoscimento facciale, espresse secondo la logica della corrispondenza, tra un input e un output, a un *token*, nella *face verification*, o a un *type*, nella *face identification*, in entrambi i casi mediante una regola che guida l'associazione, non nascono naturalmente con la *computer vision* ma si allineano con la duplice tendenza di lettura del volto che nel corso dei secoli ha caratterizzato gli studi fisiognomici (Magli 1995).

Magli, in particolare, iscrive le procedure di riconoscimento all'interno di quella tradizione fisiognomica che esegue una "minuziosa opera di pertinentizzazione del corpo" (Magli 1995, p. 32), sottolineando la natura di "impresa semiotica" di tale operazione. Estendendo il suo ragionamento, l'etichettatura delle emozioni attraverso la raccolta e la comparazione di tratti del volto operata da uno dei primi e più noti algoritmi AFR come Japanese Female Facial Expression (JAFFE) non dista molto dalle classificazioni di Charles Le Brun, teorico della patemica del volto della seconda metà del Seicento.

Le procedure di identificazione, invece, che in quanto tali portano all'"appercezione di un volto in quanto unico e irripetibile" (Magli 1995, p. 19), si riconducono alla tradizione che vede nella certificazione dell'identità di un singolo individuo un vero e proprio strumento di controllo sociale mediante la presa sul corpo.

Rispetto alle procedure di schedatura antropometrica perfezionate da Auguste Bertillon (Sekula 1986; Magli 1995; Leone 2020), i nuovi algoritmi per la *face verification* ricalibrano nuovamente la questione della sorveglianza, secondo gli attuali *surveillance studies*: è infatti un occhio inorganico, macchinico, e non una specifica figura di controllo, il solo a poter dare un'identità a immagini che sono di fatto "disembodies identities" (Gates 2011), come quelle delle telecamere di sorveglianza.

Il funzionamento della *face verification* meriterebbe di essere indagato da vicino, ma per ora, come accennato, vorremmo concentrarci sulla *face recognition* propriamente detta, all'interno della quale ricadono la maggior parte delle varietà di AFR.

A seconda delle classificazioni che vengono rese pertinenti, molti sono i modi di comparare una data immagine di input con altre immagini all'interno di un dato database, allo scopo di ottenere, come output, una etichettatura, per così dire, dell'immagine di input, in funzione di una o più categorie.

Anche in questi casi si tratta di procedure essenzialmente invisibili delle quali esperiamo solamente i risultati, come quando cerchiamo per parole chiave nella gallery del nostro smartphone o usufruiamo della ricerca per immagini sui motori di ricerca.

È difficile, insomma, arrivare a intravedere le dinamiche che consentono a una macchina di vedere e di restituirci il risultato della sua

visione — più o meno come, rimarcava Eco (1997), Marr e Nishishara (1978), nel loro lavoro sulla simulazione computerizzata dei processi percettivi, postulano una corrispondenza tra input e output (così come tra realtà e rappresentazione) ma senza indagare il perché tale corrispondenza funzioni.

Proprio per provare a capire meglio quali siano gli impliciti della *face recognition* può essere interessante guardare alle cosiddette *training images*, ovvero i grandi dataset con immagini e relative istruzioni di lettura che consentono a determinati algoritmi di elaborare input e trasformarli negli output di cui così automaticamente fruiamo. Un buon punto di partenza sono i lavori di Trevor Paglen, che da anni, sulla scena dell'arte contemporanea, si interroga sul lavoro tra arte e tecnologia.

2. *ImageNet Roulette* e *From "Apple" to "Anomaly"*: vedere come vedono le macchine

ImageNet Roulette (2019) è uno dei progetti di Trevor Paglen che sono confluiti in *Training Humans*, mostra da lui curata con Kate Crawford per Fondazione Prada Osservatorio (Milano, 12 settembre 2019–24 febbraio 2020). Come suggerisce il nome, si tratta dell'esplorazione casuale del database di immagini ImageNet.

ImageNet, in realtà, è molto più di un semplice database. Si tratta di un progetto di ricerca in collaborazione tra le università di Princeton e Stanford: nello specifico, di una immensa raccolta di immagini, prelevate da Internet, per esempio dalle immagini senza copyright su Flickr o pubblicate sui social, con l'obiettivo di "to map out the entire world of objects", nelle parole della responsabile scientifica e cofondatrice del progetto, Fei–Fei Li (cit. in Crawford–Paglen 2019).

Il dataset di ImageNet, composto da 15 milioni di immagini descritte ed etichettate da più di 50000 *workers* di 167 paesi e gratuitamente disponibile sin dal 2009 in rete, è così diventato il più grande *training set* per il *machine learning* basato sulle reti neurali convoluzionali (CNN), una particolare classe di algoritmi di apprendimento automatico, capaci ovvero di apprendere e di autoimplementarsi se "nutriti" al contempo da grandi moli di materiali e da istruzioni su come organizzarli.

Più dettagliatamente, le immagini prese dalla rete funzionano da input, e la macchina ImageNet, un vero e proprio ibrido di infrastruttura tecnologica e competenze umane (quelle dei *workers*), restituisce come output le stesse immagini ma etichettate una a una e classificate in *training set* più ristretti, raggruppati per categorie o sottocategorie (sono oltre 20000 le categorie implementate).

Output finale di ImageNet sono dunque le *training images* classificate, descritte e taggate che confluiscono nel database disponibile sul sito (https://www.image–net.org/), pronte per costituire così il cuore del processo di visione macchinica che sta alla base della *face recognition*: sono infatti lo strumento principale (il modello mentale, per dirla con Eco 1997) che consente alle reti neurali la trasformazione di ulteriori immagini di input in immagini di output, che sono, come si vedrà, degli oggetti di natura complessa, così come sono decodificati dallo sguardo artificiale.

Un'immagine letta dalle CNN non è più, infatti, solo un insieme di tratti visivi su una superficie bidimensionale, ma ha già ricevuto, grazie al confronto con *training images* che presentano pattern analoghi, un'interpretazione, condensata in una descrizione in tag che potranno a loro volta nutrire nuovi e futuri riconoscimenti di immagini.

Un esempio efficace è quello che fornisce Fei–Fei (2015), a partire dall'immagine di un bambino alle prese con una torta, immagine che, per poter essere "vista" dagli algoritmi di *machine learning*, viene segmentata in diverse aree, a seconda delle salienze evidenziate dalle categorie presenti e attive: così l'area della torta viene identificata come "cake", la sedia da giardino come "high chair", con una classificazione più dettagliata, e la figura in piedi del bambino che guarda la torta viene ricondotta all'azione di "standing on".

Come sintetizzano Mane e Shah (2019, p. 281) le reti neurali imparano a riconoscere le immagini solo se hanno a disposizione grandi database di immagini di *training* su cui esercitarsi; ed è solo grazie a questa prima fase che si sviluppa un "trained CNN model". Quello che però si tende a dare per scontato è che nonostante le reti neurali procedano individuando autonomamente, a partire dalle immagini di cui preliminarmente si nutrono, pattern e ricorrenze, non possano dare senso a ciò che registrano (non possano effettivamente "vedere") senza uno schema che sia in grado

di dire loro cosa vedono. La costruzione di questo modello mentale riposa dunque sulle associazioni categoriali depositate nelle *training images*.

Detto altrimenti, la costruzione delle *training images*, da parte dei *workers* che le processano, è il sommerso dell'*image recognition*: e, andando ancora più in profondità, senza una previa classificazione che metta a fuoco cosa è pertinente che l'algoritmo rilevi relativamente a ogni categoria, nessun riconoscimento sarebbe possibile.

Riguardo alle categorie adottate per le immagini, occorre specificare che ImageNet, sottolineano Crawford–Paglen (2019), si basa sulla struttura semantica di WordNet, un database lessicale risalente agli anni '80. La tassonomia è organizzata secondo una struttura annidata di *synset*, ovvero di sinonimi cognitivi. Ogni *synset* rappresenta un concetto distinto, con i sinonimi raggruppati, e i vari *synset* sono organizzati in una gerarchia annidata, da concetti generali a concetti più specifici. Per esempio, il concetto di *sedia* è nidificato come *artefatto>arredo>mobili>sedile>sedia*.

Mentre WordNet tenta di organizzare l'intera lingua inglese, ImageNet è limitato ai soli nomi e nella sua gerarchia ogni concetto è organizzato in una delle nove categorie di primo livello: pianta, formazione geologica, oggetto naturale, sport, artefatto, fungo, persona, animale e varie. Quello che si scopre analizzandone la struttura, tuttavia, è che talvolta si basa su categorie più o meno scientifiche (come per la classificazione degli organismi viventi, ispirata alla nomenclatura binomia di Linneo, per il genere e per la specie), ma in altri casi le classificazioni applicate sono piuttosto rozze e intuitive, quando non basate su veri e propri preconcetti.

È il caso della categoria *Person, individual, someone, somebody, mortal, soul*, su cui non a caso si esercita la critica di Paglen. Sotto la categoria *Person* (fig. 1) figurano categorizzazioni etniche, o presunte tali (*Slav*), attributi caratteriali o su stati emotivi (*Introvert*, *Optimist*), veri e propri giudizi sommari (*Creditor*, *Anti–American*, *Nonperson*, ma anche *Loser*, come in fig. 2).

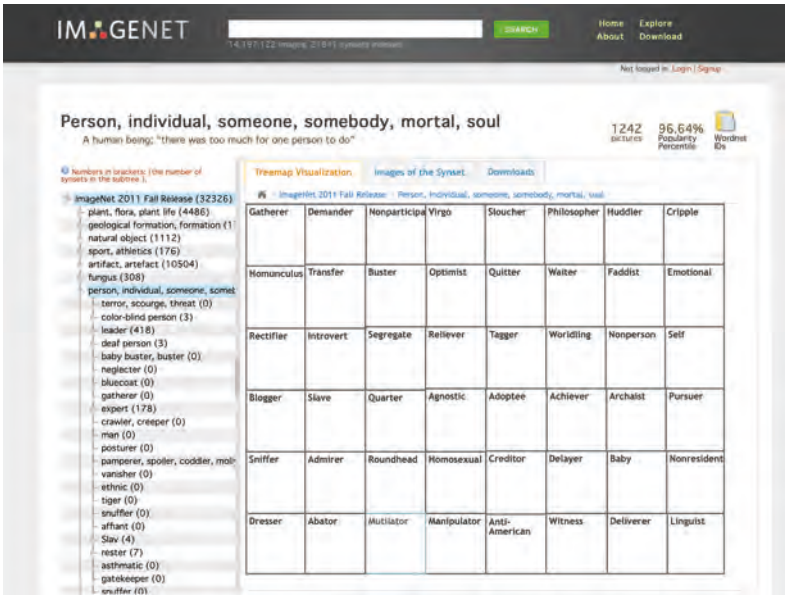

Figura 1. Alcune sottocategorie nella categoria *Person* su ImageNet, 2011 Fall Release (screenshot da Archive.org).

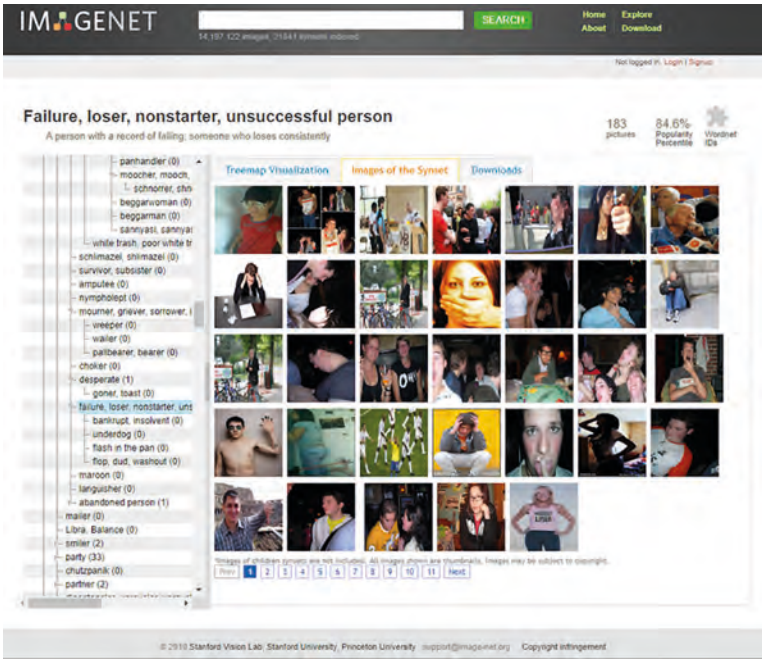

Figura 2. La sottocategoria *Failure* nella categoria *Person* su ImageNet (cit. in Crawford–Paglen 2019).

Questa tassonomia, veicolata alle CNN attraverso le immagini di *training*, è la struttura che si fa carico di filtrare l'input visivo e di restituirlo dotato di un riconoscimento: tutto questo, però, è invisibile agli utenti che ricadono nel raggio d'azione degli algoritmi per la *face recognition*. A essere visibile, dell'operato di tali *mindless agency*, sarà solo il risultato della loro applicazione.

ImageNet Roulette esibisce, invece, con una specifica interfaccia, la struttura categoriale implicita nel funzionamento per *training images*.

Il progetto (https://imagenet–roulette.paglen.com/, ora non più accessibile se non in alcuni screenshot in Crawford–Paglen 2019) per poter funzionare richiedeva all'utente il caricamento di un'immagine o un nuovo scatto. Immediatamente dopo era in grado di restituire l'output finale, filtrato alla luce delle categorie di ImageNet: la stessa immagine con il risultato in termini di *face recognitio*n, ovvero l'area del volto circoscritta da un contorno verde con applicata un'etichetta con la "lettura" del volto da parte del sistema.

I risultati sono evidenti, e notevoli, come nelle figure 3–5, prese da post di Twitter in cui gli utenti commentano la loro esperienza con *ImageNet Roulette*.

L'etichettatura operata dai *workers*, per produrre le *training images*, si applica su generiche immagini prese dalla rete, vere e proprie "disembodied identities" (Gates 2011), perché estrapolate senza più alcun riferimento al soggetto originario; quello che invece fa esplodere l'indignazione dei visitatori di *Training Humans* alle prese con ImageNet è vedere la *propria* immagine, ben ancorata alla propria identità, data in pasto a categorie che si mostrano subito per quello che sono.

Detto altrimenti, quello che astrattamente, sulle immagini del database, è percepito come classificazione, nel progetto di Paglen mostra immediatamente la sua natura di giudizio.

A ben vedere, diversi sono gli impliciti a proposito della *mind* che muove la *face recognition* portati alla luce dall'installazione di Paglen: innanzitutto, l'opacità delle categorie di classificazione, su *Person* essenzialmente declinate non come descrizioni ma come prese di posizione, e il fatto che ogni lettura basata su queste classificazioni possa produrre la cristallizzazione di nuovi stereotipi (che si autoalimentano proprio attraverso l'autoapprendimento delle reti neurali).

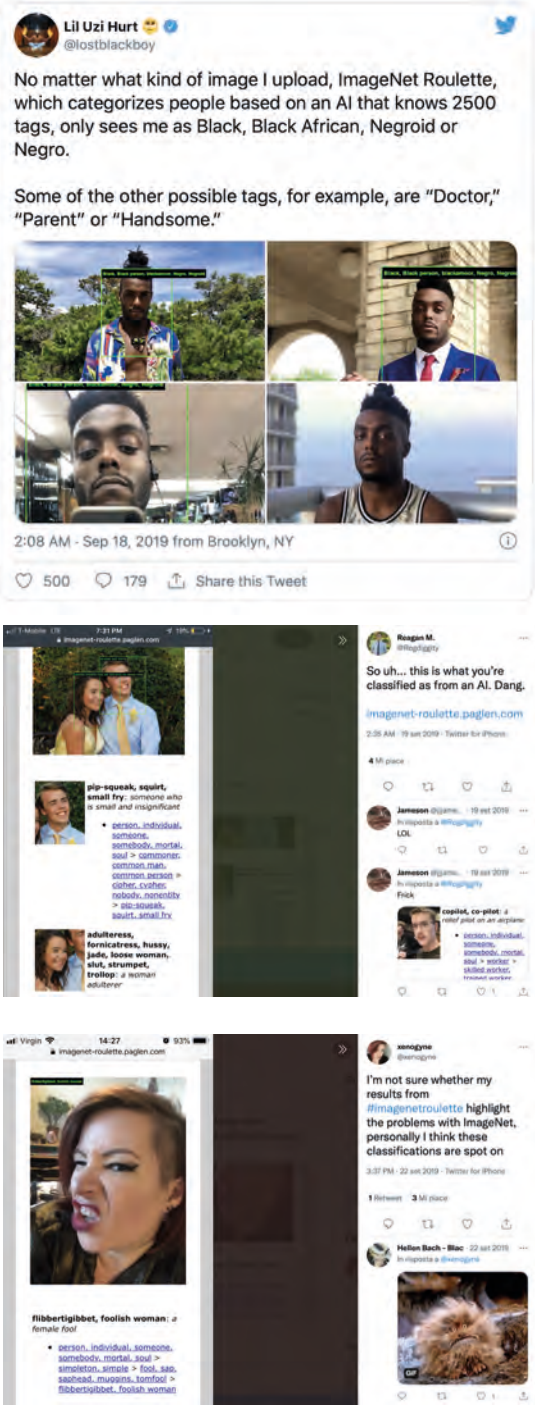

Figure 3–5. Alcuni post su Twitter su *ImageNet Roulette*.

Altro implicito non meno rilevante, il fatto che l'AFR basata sulle CNN implichi la trasformazione di una visione situata in una visione assoluta: è evidente in fig. 3, dove la prevalenza della sola etichetta *black* su ogni scatto di @lostblackboy, a prescindere dalle differenze di stile e di contesto dell'immagine, ha come risultato quello di marcare il nero, come colore della pelle, come diverso dallo standard — e mostra, in controluce, lo sguardo "bianco" e "standard" del sistema da cui ha origine. Sguardo "bianco" che quando si esercita sulla foto di @ Rogdiggity (fig. 4) non lo marca per il colore della pelle, ma attraverso altri tag. Discorso a parte meriterebbero poi le etichette vistosamente misogine (vedi fig. 4, la ragazza in basso).

Ecco che l'AFR emerge anche come terreno di scontro e di definizione dell'alterità oltre che dell'identità, non perché crei autonomamente nuove categorie ma perché fornisce un nuovo, ancora più pericoloso perché invisibile, campo di applicazione delle categorie dell'alterità, che da sempre si fondano su una specifica polarità, volta a identificare da una parte lo standard e di converso la deviazione da esso. Se si parla di categorie visive dell'alterità questo meccanismo di doppia messa a fuoco è ancora più evidente, come mostra Stoichita (2019) portando a esempio la cosiddetta Negra di padre Gumilla, affetta da una sorta di vitiligine, la cui pelle è ostinatamente descritta non come nera con problemi di pigmentazione ma come ricoperta da una sorta di "mantelletta nera" e da "guanti neri" (Le Cat, 1765, cit. in Stoichita 2019) quasi che, dal punto di vista del bianco esploratore, le parti nere non potessero essere viste che come accessorie, un di più rispetto alla "normale" pelle bianca.

Un altro implicito che *ImageNet Roulette* porta alla luce è la trasformazione delle espressioni mutevoli del volto delle immagini di input in classificazioni permanenti nei risultati di output, come per esempio in fig. 5. Il diverso intento classificatorio che ha da sempre orientato e distinto *fisiognomica* (come studio delle strutture del volto) e *patognomica* (come studio delle passioni espresse dal volto) si fonde insomma nelle bizzarre categorizzazioni applicate dai *workers*, che riassumono in tag sommari (dalle connotazioni spesso decisamente misogine o razziste) espressioni e atteggiamenti del volto.

La riflessione di *ImageNet Roulette* trova, ci sembra, un ulteriore approfondimento in un altro progetto di Paglen, *From "Apple" to*

"Anomaly" (2019), installazione per il Barbican Center. Il materiale di partenza è sempre ImageNet, ma lo scopo del progetto è differente. Al posto dell'esplicitazione delle strutture classificanti, c'è l'esibizione simultanea di una porzione di materiale, corrispondente a più di 200 categorie. Le *training images* sono riprodotte insieme e allineate senza soluzione di continuità su un lungo muro del centro espositivo. L'esplorazione va da *apple*, categoria com'è prevedibile molto facile da segmentare tassonomicamente, ad *Anomaly*, l'etichetta generica che ImageNet attribuisce alle immagini che non riesce a classificare, ovvero che ricadono del tutto fuori dalle maglie del sistema. Il che però porta a un paradosso: anche la categoria *Anomaly*, che denuncia l'impossibilità di una classificazione, è una categoria che opera una classificazione.

Per costruzione, per taglio, per orientamento, la macchina non riesce a vedere queste immagini, perché non trova dei match tra immagine di input e immagini del *training set* lungo nessuna delle categorie previste; ma l'inclassificabilità in termini algoritmici confina le immagini "devianti" dal sistema in una categoria a parte, di fatto stigmatizzandole.

La visione artificiale, la visione aggregata della macchina, in ultima analisi, individua un suo personale visibile circoscrivendo però anche quello che è per lei non visibile, ovvero che non rientra tra le sue categorie. Una discriminazione per difetto di visione (cfr. anche Buolamwini, Gebru 2018), che potremmo definire "passiva", in contrapposizione alle pratiche di discriminazione "attiva", per marcatura, che allineano determinati input visivi in categorie biometriche molto specifiche, per esempio implementando negli algoritmi per la scansione dei tratti facciali le caratteristiche somatiche di un dato gruppo etnico come nel noto caso degli uiguri in Cina (Mozur 2019), operando per *social sorting* (Lyon 2003).

3. Da *training images* a *operational images*: guardare alle immagini come dati

Sia *ImageNet Roulette* che *From "Apple" to "Anomaly"*, in ultima analisi, offrono all'osservatore una prospettiva privilegiata per riflettere sugli impliciti della *face recognition*, mostrando molto chiaramente la

necessità di pensare alla visione artificiale come a una visione che potremmo definire *informazionale*, proprio perché mette in atto delle dinamiche di classificazione a partire dalla specificità del singolo volto.

La *face recognition*, infatti, riduce la complessità dell'immagine del volto a un insieme di dati, attraverso l'imposizione di una data prospettiva di visione che passa per la messa a punto di categorie (e di modelli descrittivi) e l'applicazione di tali categorie, proprio come nell'*information design* sia la trasformazione dei *raw data* in informazione strutturata che la scelta di una specifica modalità di visualizzazione per l'informazione sono oggetto di pratiche di traduzione (Manchia 2020).

Detto altrimenti, fondamentalmente non esistono immagini, così come le intendiamo noi, per le reti neurali, ma piuttosto dati che possono essere organizzati in informazione, ovvero che diventano leggibili solo grazie all'intervento di queste strutture "pensanti" che li lavorano, tra le maglie della visione artificiale.

Per questo motivo può essere interessante ripensare alle *training images* non più come immagini ma come *information*, per esempio sulla scia del lavoro di ricerca di Harun Farocki sulle immagini di sorveglianza, su cui anche Paglen ha lungamente riflettuto (cfr. per esempio Paglen 2014).

In particolare, proponiamo di riprendere il concetto di *operational images* che Farocki (2005) conia per riferirsi alle immagini che le telecamere di sorveglianza producono, nel senso di "images that are not simply meant to reproduce something, but instead are part of an operation." Il girato cieco e automatico dei dispositivi a circuito chiuso diventa infatti per noi una serie di immagini, dotata di un valore e di un senso, solo all'interno di una narrazione orientata, come quella di Farocki in *Eye/Machine* (2001–2003) o in *War at Distance* (2003).

Fuori dal montaggio che ci fa conoscere quelle immagini come immagini, le *operational images* sono infatti puri pezzi di operatività per il sistema che li produce, elementi che il sistema produce per se stesso e nei quali noi potremmo anche non avere alcuna parte.

Ci sembra che si possano fare delle considerazioni simili anche sulle *training images*. Quello che forse ci colpisce di più, dunque, in operazioni come quelle di Paglen, è che mostrano con chiarezza l'esistenza di un mondo di immagini non destinate a noi, "invisible images" (Paglen

2019), appunto, che delle immagini hanno l'apparenza, per il nostro sguardo analogico, mentre per la visione artificiale non sono altro che un input di partenza, il campo di applicazione di algoritmi volti all'ottenimento di specifici risultati.

4. Epilogo. Il prossimo futuro della visione artificiale

In chiusura, occorre aggiungere che, nei mesi che sono seguiti allo scandalo sollevato da *ImageNet Roulette*, a opera del gruppo di ricercatori capitanato da Fei–Fei Li ci sono stati degli ulteriori, importanti sviluppi.

Nello specifico, in Yang *et al.* (2020) sono stati riscontrati tre principali problemi da correggere nel subtree *Person*: lo "stagnant concept vocabulary of WordNet", che si propone di correggere attraverso l'individuazione (e l'eliminazione) delle categorie *unsafe* perché *offensive* o *sensitive*, ovvero che sono "not inherently offensive but may cause offense when applied inappropriately"; la carente rappresentazione delle minoranze nel corpus di immagini di ImageNet, risolvibile implementando nel database immagini che rappresentino uno spettro più ampio di casistiche rispetto al genere, al colore della pelle o all'età, correggendo così i *bias* che portano di fatto alla cancellazione di alcune casistiche; e, infine, l'impossibilità di individuare un'equivalenza visiva per ogni *synset* di WordNet.

Più in dettaglio, molte delle distorsioni rese evidenti da *ImageNet Roulette* vengono ricondotte all'impossibilità di avere sempre una corrispondenza efficace tra *synset* e immagini, ovvero dalla *non–imageability* di alcuni concetti e sulla conseguente necessità di introdurre un *imageability score* che tagli fuori, dall'azione dell'AFR, i concetti che ottengono valori più bassi in sede sperimentale.

Viene però da chiedersi, in una prospettiva che non è più quella della *computer vision* ma degli studi visivi, tra semiotica, teoria delle immagini e *visual culture studies*, che cosa implichi cercare di vagliare, come i ricercatori di ImageNet fanno, le categorie chiave per la *computer vision* sulla base di un *imageability score* che misuri l'immediatezza attraverso la quale un concetto può essere tradotto in un'immagine. Non sarebbe piuttosto utile riflettere, a monte, sulla diversità intrinseca di

un concetto come quello di *apple* (per riprendere l'esempio di Paglen 2019), decisamente più *nouny* (Lakoff 1987), rispetto a un concetto astratto, perché può rifarsi a una configurazione data di tratti visivi che corrispondono, a partire da una specifica griglia di lettura, a dati tratti del mondo naturale (Greimas 1984)?

Yang *et al.* (2020) ammette poi l'esistenza di categorie "imageable, hard to classify", anche nel caso di concetti apparentemente molto semplici come *basketball player*.

Ma chi è davvero un *basketball player* per le CNN? È di default un *giocatore di basket*, o soltanto una *persona che gioca a basket*? Nella stessa categoria non possono che convivere un'accezione ristretta e un'accezione più allargata del termine: certamente il *setting* in cui l'azione si svolge, che enciclopedicamente indirizza la lettura (Eco 1975), e le istruzioni per il riconoscimento che si basano su precedenti esperienze percettive (Eco 1997) orientano la nostra visione, ma come rendere tutto questo accessibile agli algoritmi?

Esiste il rischio concreto, insomma, che *computer vision* e *AFR* vedano il mondo solo a misura delle categorie che contribuiscono a incasellarlo, come nelle pericolose derive nel controllo sugli uiguri e altre minoranze (cfr. anche Manchia 2021).

Vedere, insomma, si dice in molti modi, per gli umani ma anche per le macchine: e sarebbe interessante continuare a riflettere da vicino sulla grande distanza che separa questi due sguardi, che di fatto individuano due mondi del visibile non immediatamente sovrapponibili tra di loro.

Riferimenti bibliografici

Arcagni S. (2018) *L'occhio della macchina*, Einaudi, Torino.

Buolamwini J., Gebru T. (2018) *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*. Proceedings of the 1st Conference on Fairness, Accountability and Transparency, in «Proceedings of Machine Learning Research», 81:77–91, https://proceedings.mlr.press/v81/buolamwini18a.html.

Crawford K., Paglen T. (2019) *Excavating AI: The Politics of Training Sets for Machine Learning*, September 19, 2019, https://excavating.ai.

Eco U. (1975) *Trattato di semiotica generale*, Bompiani, Milano.

Eco U. (1997) *Kant e l'ornitorinco*, Bompiani, Milano.

Farocki H. (2005) *Der Krieg findet immer einen Ausweg*, in *Cinema 50. Essay*, Schüren Verlag, Marburg: 21–33 (trad. eng. *War Always Finds a Way*, in C. Pontbriand (a cura di), *HF/RG — Harun Farocki/Rodney Graham*, Jeu de Paume/Blackjack Editions, Paris, 2009, 107.

Fei–Fei L. (2015) *How we're teaching computers to understand pictures*, TED Talk, https://www.ted.com/talks/fei_fei_li_how_we_re_teaching_computers_to_understand_pictures.

Floridi L. (2019) *What the near future of artificial intelligence could be*, «Philosophy & Technology», 32(1): 1–15.

Floridi L. (2020) *Artificial Intelligence as a Public Service: Learning from Amsterdam and Helsinki*, «Philos. Technol.», 33: 541–546.

Gates K.A. (2011) *Our Biometric Future. Facial Recognition Technology and the Culture of Surveillance*, New York University Press, New York.

Greimas A.J. (1984) *Sémiotique figurative et sémiotique plastique*, «Actes sémiotiques. Documents», n. 60 (trad. it. *Semiotica figurativa e semiotica plastica*, in P. Fabbri, G. Marrone (a cura di), *Semiotica in nuce II. Teoria del discorso*, Meltemi, 2001, pp. 196–210).

Lakoff G. (1987) *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*, University of Chicago Press, Chicago (trad. it. *Donne, fuoco e cose pericolose*, La Nuova Italia Editrice, Scandicci, 1999).

Leone M. (2018) *The Semiotics of the Face in the Digital Era*, «Perspectives», 17: 27–29.

Leone M. (2020) *From Fingers to Faces: Visual Semiotics and Digital Forensics*. «International Journal for the Semiotics of Law«, online, https://doi.org/10.1007/s11196–020–09766–x.

Lyon D. (2003) *Surveillance as Social Sorting: Computer Codes and Mobile Bodies*, in D. Lyon (a cura di) *Surveillance as Social Sorting: Privacy, Risk and Digital Discrimination*, Routledge, London, 13–30.

Lyon D. (2018) *The Culture of Surveillance: Watching as a Way of Life*, Polity Press, London (trad. it. *La cultura della sorveglianza. Come la società del controllo ci ha reso tutti controllori*, Luiss University Press, Roma, 2020)

Magli P. (1995) *Il volto e l'anima. Fisiognomica e passioni*, Bompiani, Milano.

Manchia V. (2020) *Il discorso dei dati. Note semiotiche sulla visualizzazione delle informazioni*, FrancoAngeli, Milano.

Manchia V. (2021) *Dati mancanti, dati mancati. Cancel culture, data bias e data gap nell'era dei big data*, «Filosofi(e)Semiotiche», vol. 8, n. 1, pp. 62–73.

Mane S., Shah G. (2019) *Facial Recognition, Expression Recognition, and Gender Identification*, in V.E. Balas *et al.* (a cura di), *Data Management, Analytics and Innovation*, «Advances in Intelligent Systems and Computing», 808, v. 808. Springer, Singapore, https://doi.org/10.1007/978–981–13–1402–5_21.

Manovich L. (2018) *AI Aesthetics*, Strelka Press, Moscow (trad. it. *L'estetica dell'intelligenza artificiale*, Luca Sossella editore, Roma, 2020).

Marr D. (1978) *Computer Vision System*, Academic Press, New York.

Marr D. (1982) *Vision. A Computational Investigation into the Human Representation and Processing of Visual Information*, MIT Press, Cambridge, Mass., 2010.

Marr D., Nishishara H.K. (1978) *Representation and recognition of the spatial organization of three–dimensional shapes*, «Proc. R. Soc. Lond.», B.200269–294, http://doi.org/10.1098/rspb.1978.0020.

McCarthy J., Minsky M.L., Rochester N., Shannon C.E. (1955) *A proposal for the Dartmouth summer research project on artificial intelligence*, «AI Magazine"», August 31, 1955. 27(4), 12, 2006.

Metz C. (2019) *"Nerd", "Nonsmoker", "Wrongdoer": How Might A.I. Label You?*, «New York Times«, 20 settembre 2019, https://www.nytimes.com/2019/09/20/arts/design/imagenet–trevor–paglen–ai–facial–recognition.html.

Mozur P. (2019) *How China Is Using AI to Profile a Minority*, «The New York Times», 14 aprile 2019, https://www.nytimes.com/2019/04/14/technology/china–surveillance–artificial–intelligence–racial–profiling.html.

Paglen T. (2004) *Operational Images*, «e–flux», 59, http://e–flux.com/journal/59/61130/operational–images.

Paglen T. (2019) *Invisible Images: Your Pictures Are Looking at You*, «Architectural Design», 89: 22–27.

Sekula A. (1986) *The Body and the Archive*, *October*, 39: 3–64.

Shannon C.E. (1950) *Programming a Computer for Playing Chess*, «Philosophical Magazine», 7, vol. 41, n. 314, https://users.dcc.uchile.cl/~cgutierr/cursos/IA/shannon.txt.

Stoichita V. (2019) *L'immagine dell'Altro. Neri, giudei, musulmani e gitani nella pittura occidentale dell'Età moderna*, La Casa Usher, Firenze.

Turing A.M. (1950) *Computing Machinery and Intelligence*, «Mind», 59: 433–460.

Wiener N. (1965) *Cybernetics, or Control and Communication in the Animal and the Machine*, MIT Press, Cambridge, Mass. (trad. it. *La Cibernetica. Controllo e comunicazione nell'animale e nella macchina*, il Saggiatore, Milano, 1968).

Yang K., Qinami K., Fei–Fei L., Deng J., Russakovsky O. (2020) *Towards fairer datasets: filtering and balancing the distribution of the people subtree in the ImageNet hierarchy*, «Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency», pp. 547–558, https://doi.org/10.1145/3351095.3375709.