

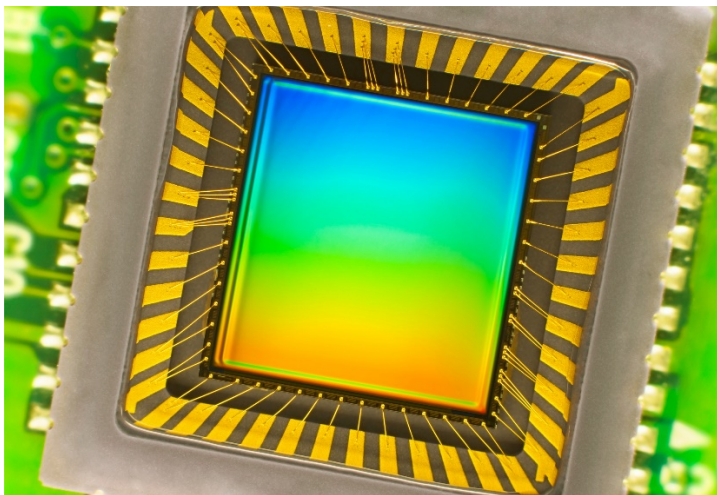
La fotografia “computazionale”

([Devin Coldewey@techcrunch](mailto:DevinColdewey@techcrunch))

Un nuovo mondo per la cattura dell'immagine

Cosa c'è in una macchina fotografica? Un obiettivo, un otturatore, una superficie sensibile alla luce e, sempre più, un insieme di algoritmi (procedure di calcolo) altamente sofisticati. Mentre i componenti fisici stanno ancora migliorando a piccoli passi, Google, Samsung e Apple stanno investendo sempre di più (e lo dimostrano) in miglioramenti apportati interamente dagli algoritmi di calcolo implementati nei loro smartphones. La fotografia computazionale è diventato così il loro unico vero campo di battaglia. La ragione di questo cambiamento è piuttosto semplice: le macchine fotografiche non possono migliorare molto di più di quanto non lo siano in questo momento, o almeno non senza alcuni cambiamenti piuttosto estremi nel loro modo di lavorare. Ecco come i produttori di smartphones si sono scontrati con questo muro nella fotografia e come lo hanno scavalcato.

Non ci sono abbastanza secchielli (raccoglitori di luce)



Un sensore di immagine come lo si può trovare in una fotocamera digitale

I sensori delle nostre fotocamere per smartphone sono cose davvero incredibili. Il lavoro svolto da produttori del calibro di Sony, OmniVision, Samsung e altri per progettare e fabbricare chip piccoli ma sensibili e versatili è davvero piuttosto strabiliante. Per un fotografo che ha seguito l'evoluzione della fotografia digitale fin dai primi giorni, il livello di qualità offerto da questi sensori microscopici è a dir poco stupefacente.

Ma per questi sensori non vale la legge di Moore, che predice un raddoppio del numero di transistor per unità di superficie di silicio ogni 18 mesi. O meglio, proprio come la legge di Moore sta ora arrivando ai limiti quantici per i processori con dimensioni dei transistor inferiori a 10 nanometri, i sensori delle fotocamere sono arrivati ai limiti fisici molto prima. Pensiamo alla luce che colpisce il sensore come se fosse la pioggia che cade su un mucchio di secchielli; puoi mettere secchielli più grandi, ma ce ne saranno meno per unità di superficie; puoi metterne di più piccoli, ma riusciranno a catturare meno gocce. Puoi renderli quadrati o scaglionarli in maniera differente o tentare altri tipi di trucchi, ma

alla fine ci sono solo tante gocce di pioggia e nessuna quantità di secchielli o di riorganizzazione dei secchielli ne può raccogliere di più.

I sensori stanno migliorando, sì, ma non solo questo ritmo è troppo lento per invogliare i consumatori a comprare nuovi telefoni anno dopo anno (immagina di provare a vendere una fotocamera migliorata solo del 3%), ma i produttori di telefoni usano spesso tipi di sensori uguali o simili, quindi i miglioramenti (come il recente passaggio alla retroilluminazione) sono condivisi tra loro. Quindi nessuno di loro può avvantaggiarsi basandosi solo sui sensori.

Forse potrebbero migliorare l'obiettivo? Non proprio. Le lenti sono arrivate a un livello di sofisticazione e perfezione tale che è difficile migliorarle, specialmente su piccola scala. Dire che lo spazio è limitato all'interno di queste fotocamere per smartphone è un eufemismo, non c'è quasi un micron quadrato da risparmiare. Le si potrebbe migliorare leggermente per aumentare il passaggio della luce e diminuire la poca distorsione che hanno ancora, ma questi sono vecchi problemi che sono stati per lo più ottimizzati.

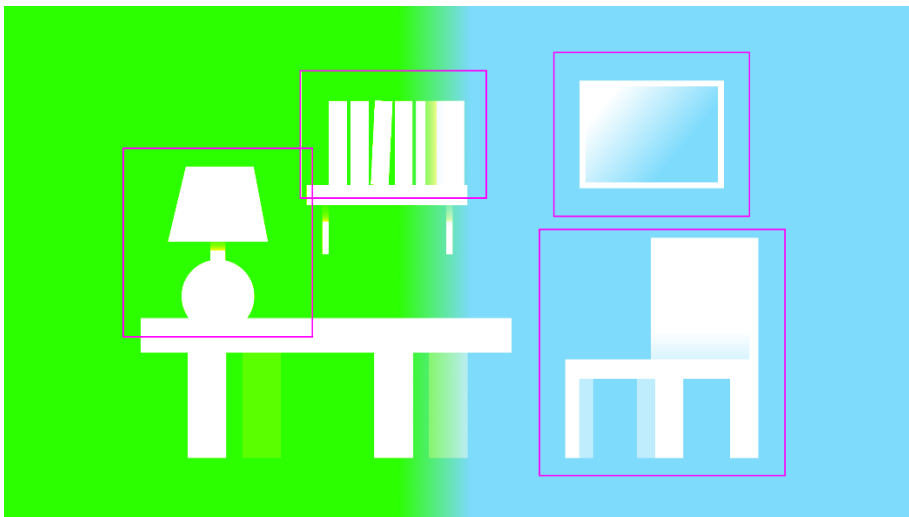
L'unico modo per raccogliere più luce sarebbe quello di aumentare la dimensione della lente, oppure A: ponendola all'esterno dello smartphone; B: spostare i componenti critici all'interno del corpo dello smartphone; C: aumentare lo spessore del telefono. Quale di queste opzioni è probabile che Apple o un altro produttore ritenga accettabile?

Era inevitabile che Apple, Samsung, Huawei e altri scegliessero ancora un'altra soluzione. Se non puoi ottenere più luce, **devi solo fare di più con la luce che hai.**

Ma la fotografia non é quasi tutta “computazionale”?

La più ampia definizione di fotografia computazionale include quasi tutte le immagini digitali. *A differenza della pellicola, anche la fotocamera digitale più basilare richiede dei calcoli per trasformare la luce che colpisce il sensore in un'immagine utilizzabile.* E i produttori di macchine fotografiche differiscono ampiamente nel modo in cui lo fanno, utilizzando diversi metodi di creazione di immagini in formato JPEG, differenti formati RAW e di elaborazione del colore.

Per molto tempo non c'è stato molto interesse per questo tipo di calcoli, principalmente per la mancanza di potenza di elaborazione. Certo, si sono creati dei filtri e dei piccoli interventi della fotocamera per migliorare il contrasto e il colore. Ma alla fine questi erano interventi trascurabili.

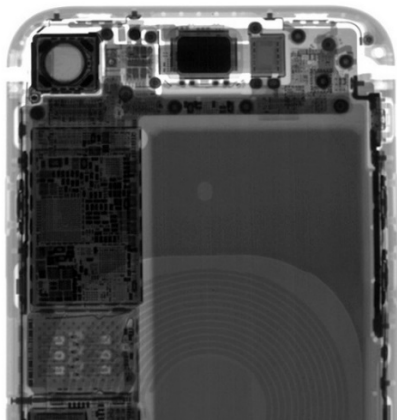


Le prime vere caratteristiche computazionali della fotografia erano probabilmente l'identificazione e il tracciamento degli oggetti ai fini dell'autofocus. Il rilevamento di volti e occhi rendeva più facile catturare persone in situazioni di illuminazione o pose complesse,

e il tracciamento degli oggetti rendeva più facile la fotografia sportiva e d'azione mantenendo il punto AF su un punto/elemento che si spostava attraverso il fotogramma. Questi sono stati i primi esempi per derivare metadati dall'immagine e di utilizzarli in modo proattivo, per migliorare quell'immagine o la successiva.

Nelle reflex digitali, la precisione e la flessibilità dell'autofocus sono caratteristiche di selezione, quindi questo caso di utilizzo iniziale aveva senso; ma al di fuori di alcuni espedienti, queste macchine fotografiche "serie" generalmente implementavano il calcolo in modo abbastanza limitato. Sensori di immagine più veloci significavano maggiore velocità di scaricamento del sensore e velocità di burst (numero di scatti sequenziali), alcuni cicli extra dedicati alla conservazione del colore e dei dettagli e così via. Le reflex digitali non venivano utilizzate per video live o realtà aumentata. E fino a poco tempo fa, lo stesso valeva per le fotocamere degli smartphone, che erano da usare come "prendi e scatta" più che agli strumenti multimediali per tutti gli usi che conosciamo oggi.

I limiti dell'imaging tradizionale



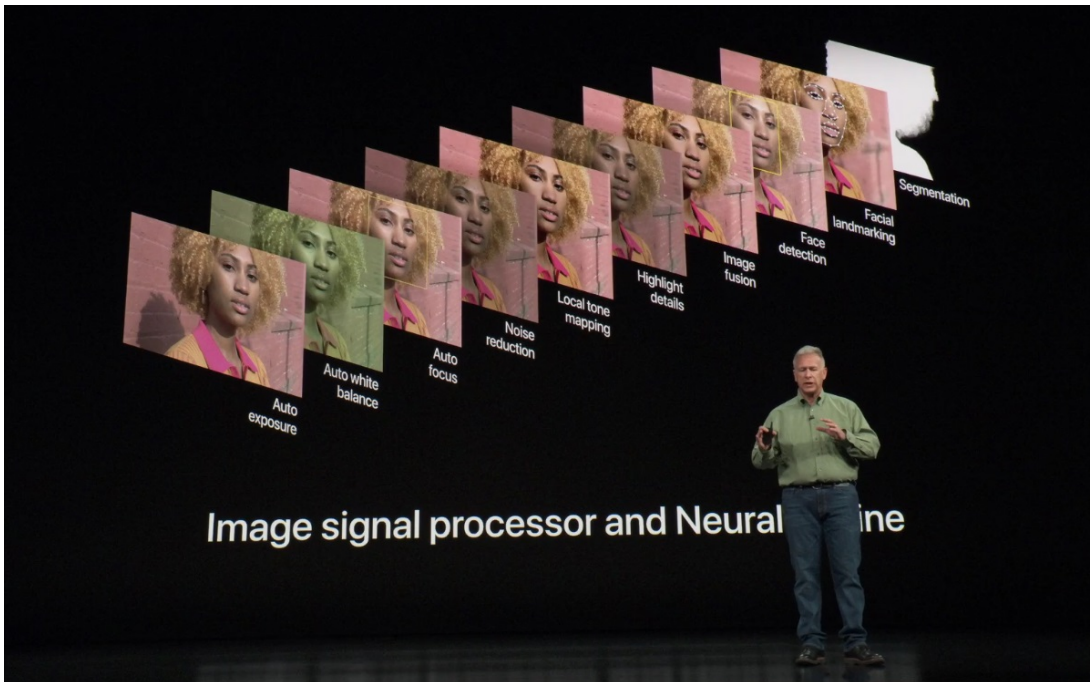
Nonostante le sperimentazioni qua e là e le anomalie occasionali, le fotocamere per smartphone sono praticamente le stesse. Devono adattarsi a pochi millimetri di profondità, il che limita la loro ottica a poche configurazioni. Allo stesso modo, la dimensione del sensore è limitata: una DSLR potrebbe utilizzare un sensore APS-C 23 di 15 millimetri di diametro, creando un'area di 345 mm²; il sensore dell'iPhone XS, probabilmente il più grande e avanzato sul mercato in questo momento, è 7 per 5,8 mm o giù di lì, per un totale di 40,6 mm².

In parole povere, sta raccogliendo un ordine di grandezza in meno di luce rispetto a una fotocamera "normale", ma si prevede che ricostruisca una scena con la stessa fedeltà, i colori e così via – e con lo stesso numero di megapixel. Un tipo di problema impossibile da risolvere.

Miglioramenti nel senso tradizionale aiutano: la stabilizzazione ottica ed elettronica, ad esempio, consente di esporre più a lungo senza sfocature, raccogliendo più luce. Ma a questi dispositivi viene ancora chiesto di trasformare la paglia in oro.

Fortunatamente, come ho detto, tutti i produttori sono praticamente sulla stessa barca. A causa dei limiti fondamentali in gioco, Apple o Samsung o ... non possono reinventare la fotocamera o inventare una struttura di obiettivi pazzesca che li metta in vantaggio rispetto alla concorrenza. Tutti hanno lo stesso punto di partenza. Perciò la competizione tra i differenti produttori nasce da quello che riusciranno a sviluppare "oltre" quella base.

Immagine come flusso



L'intuizione chiave nella fotografia computazionale è che un'immagine proveniente dal sensore di una fotocamera digitale non è un'istantanea, come generalmente si pensa. Nelle fotocamere tradizionali l'otturatore si apre e si chiude, esponendo il mezzo fotosensibile per una frazione di secondo. Non è quello che fanno le fotocamere digitali, o almeno non quello che possono fare.

Il sensore di una fotocamera è costantemente bombardato dalla luce; la pioggia cade costantemente sul campo dei secchielli, per tornare alla nostra metafora, ma quando non fai una foto, questi secchielli sono senza fondo e nessuno sta controllando il loro contenuto. Ma la pioggia cade comunque.

Per catturare un'immagine, il sistema della fotocamera sceglie un punto in cui iniziare a contare le gocce di pioggia, misurando la luce che colpisce il sensore. Quindi sceglie un punto per fermarsi. Ai fini della fotografia tradizionale, ciò consente tempi di scatto quasi arbitrariamente brevi, che non sono molto utili per i sensori minuscoli.

Perché non registrare sempre? Teoricamente si potrebbe, ma si scaricherebbe la batteria e si produrrebbe un sacco di calore. Fortunatamente, negli ultimi anni i chip di elaborazione delle immagini sono diventati abbastanza efficienti da poter, quando l'app della fotocamera è aperta, mantenere una certa durata di quel flusso, ad esempio, con una risoluzione limitata agli ultimi 60 fotogrammi. Certo, costa un po' di batteria, ma ne vale la pena.

L'accesso al flusso consente alla fotocamera di fare ogni genere di cose. Aggiunge contesto.

Il contesto può significare molte cose. Possono essere elementi fotografici come l'illuminazione e la distanza dal soggetto. Ma può anche essere movimento, oggetti, intenzione.

Un semplice esempio di contesto è ciò che viene comunemente definito HDR o immagini ad alta gamma dinamica. Questa tecnica utilizza più immagini prese di seguito con esposizioni diverse per acquisire in modo più accurato aree dell'immagine che potrebbero essere state sottoesposte o sovraesposte in una singola esposizione. Il contesto in questo caso è capire quali aree sono e come combinare in modo intelligente le immagini.

Questo può essere ottenuto con bracketing dell'esposizione, una tecnica fotografica molto antica, ma può essere eseguita immediatamente e senza preavviso se il flusso di immagini viene manipolato per produrre più intervalli di esposizione per tutto il tempo. Questo è esattamente ciò che fanno ora Google e Apple.



Qualcosa di più complesso è naturalmente la "modalità ritratto" e sfocatura dello sfondo artificiale o bokeh che sta diventando sempre più comune. Il contesto qui non è semplicemente la distanza di un volto, ma la comprensione di quali parti dell'immagine costituiscono un particolare oggetto fisico e i contorni esatti di quell'oggetto. Questo può essere derivato dal movimento nel flusso, dalla separazione stereo in più fotocamere e dai modelli di apprendimento automatico che sono stati "addestrati" per identificare e delineare le forme umane.

Queste tecniche sono possibili solo, in primo luogo, perché le immagini richieste sono state acquisite dal flusso (un anticipo nel sensore di immagine e nella velocità della RAM) e in secondo luogo, perché le aziende hanno sviluppato algoritmi altamente efficienti per eseguire questi calcoli, basatisu enormi set di dati e quantità immense di tempo di calcolo.

Ciò che è importante per queste tecniche, tuttavia, non è semplicemente che possono essere fatte, ma che una società può farlo meglio dell'altra. E questa qualità è interamente una funzione del lavoro di ingegneria del software e della supervisione artistica che li riguarda.

DxOMark ha fatto un confronto tra alcuni primi sistemi di bokeh artificiale; i risultati, tuttavia, erano in qualche modo insoddisfacenti. Era meno una questione di quale aspetto fosse migliore, ma più se fallivano o riuscivano ad applicare l'effetto. La fotografia computazionale è così giovane che è sufficiente che la funzione "esista" semplicemente per impressionare le persone. Come un cane che cammina sulle zampe posteriori, siamo stupiti che ciò avvenga.

Ma Apple ha portato avanti quello che alcuni direbbero una soluzione quasi assurdamente sovradimensionata al problema del bokeh. Non ha solo imparato a replicare l'effetto: **ha usato la potenza di calcolo che ha a disposizione per creare modelli fisici virtuali del fenomeno ottico che lo produce.** È come la differenza tra l'animazione di una palla che rimbalza e la simulazione della gravità reale e della fisica dei materiali elastici.

Perché andare così lontano? Perché Apple sa cosa sta diventando chiaro a tutti: che è assurdo preoccuparsi dei limiti della capacità computazionale. Ci sono dei limiti al modo in cui un fenomeno ottico può essere replicato se si stanno prendendo scorciatoie come la sfocatura gaussiana. Non ci sono limiti a quanto bene possa essere replicato se lo simuli al livello del fotone (singolo elemento della luce).

Allo stesso modo l'idea di combinare cinque, dieci o cento immagini in un'unica immagine HDR sembra assurda, ma la verità è che nella fotografia, più informazioni utilizzi, più il risultato migliora. Se il costo di queste acrobazie computazionali è trascurabile e i risultati misurabili, perché i nostri dispositivi non dovrebbero eseguire questi calcoli? Tra qualche anno anche loro sembreranno normali.

Se il risultato è un prodotto migliore, la potenza computazionale e le capacità ingegneristiche sono state implementate con successo; proprio come Leica o Canon potrebbero spendere milioni per ottenere miglioramenti delle prestazioni da un sistema ottico stabile come un obiettivo zoom da \$ 2,000, Apple e altri stanno spendendo soldi dove possono creare valore: non in vetro, ma in silicio.

Visione doppia

Una tendenza che potrebbe sembrare in conflitto con la fotografia computazionale che ho descritto è l'avvento di sistemi che comprendono più fotocamere.

Questa tecnica non aggiunge più luce al sensore, cosa che sarebbe proibitivamente complessa e costosa in ottica e probabilmente non funzionerebbe comunque. Ma se si riesce a liberare un po' di spazio nel senso della lunghezza (piuttosto che in profondità, che è poco pratico) si può posizionare una fotocamera completamente separata dalla prima e che catturi immagini estremamente simili.



Un mock-up di come potrebbe apparire una linea di iPhone a colori

Ma nessuno in realtà vuole riprendere due immagini contemporaneamente, con fotocamere distanti una frazione di centimetro.

Queste due fotocamere funzionano indipendentemente (come grandangolo e zoom) o una viene utilizzata per aumentare l'altra, formando un singolo sistema con più ingressi. Il fatto è che prendere i dati da una fotocamera e utilizzarli per migliorare i dati di un'altra è - lo indovinate - estremamente intensivo dal punto di vista computazionale. È come il problema HDR di esposizioni multiple, tranne che molto più complesso in quanto le immagini non vengono scattate con lo stesso obiettivo e sensore. Può essere ottimizzato, ma questo non lo rende facile.

Quindi, anche se l'aggiunta di una seconda fotocamera è davvero un modo per migliorare il sistema fotografico con mezzi fisici, la possibilità esiste solo grazie alla fotografia computazionale. Ed è la qualità di quell'immagine computazionale che si traduce in una fotografia migliore o meno. La fotocamera Light con i suoi 16 sensori e obiettivi è un esempio di uno sforzo ambizioso che semplicemente non ha prodotto immagini migliori, anche se utilizzava tecniche di fotografia computazionale consolidate per raccogliere e ottenere una raccolta ancora più ampia di immagini.

Luce e algoritmi

Il futuro della fotografia è computazionale, non ottico. Questo è un enorme cambiamento di paradigma e ogni azienda che produce o utilizza fotocamere lo sta affrontando. Ci saranno ripercussioni nelle fotocamere tradizionali come le reflex (che stanno rapidamente cedendo il passo ai sistemi mirrorless), nei telefoni, nei dispositivi integrati e ovunque la luce viene catturata e trasformata in immagini.

A volte questo significa che le fotocamere di cui parliamo saranno più o meno le stesse dell'anno scorso, per quanto riguarda i conteggi di megapixel, le gamme ISO, i numeri f e così via. Va bene. Con alcune eccezioni, queste sono diventate buone quanto possiamo ragionevolmente aspettarci che siano: il vetro non sta diventando più chiaro e la nostra visione non sta diventando più acuta. Il modo in cui la luce si muove attraverso i nostri dispositivi e i nostri occhi non cambierà molto.

Ciò che questi dispositivi fanno con quella luce, tuttavia, sta cambiando a un ritmo incredibile. Ciò produrrà caratteristiche che suonano ridicole o pseudoscienza o ... batterie scariche. Va bene anche questo. Proprio come abbiamo sperimentato con altre parti della fotocamera per il secolo scorso e li abbiamo portati a diversi livelli di perfezione, siamo passati a una nuova "parte" non fisica che tuttavia ha un effetto molto importante sulla qualità e persino sulla possibilità delle immagini che riprenderemo.