

Ricostruzione Agnostica di Immagini Compresse con Generative Adversarial Networks

Leonardo Galteri, Lorenzo Seidenari, Marco Bertini, Alberto Del Bimbo
Media Integration and Communication Center, Università degli Studi di Firenze
{name.surname}@unifi.it

Abstract

La compressione delle immagini è un processo necessario in molte circostanze. Tuttavia, l'utilizzo di un algoritmo di compressione con perdita di dati comporta la generazione di artefatti nelle immagini. Di conseguenza, la percezione umana dell'immagine viene alterata, risultando meno gradevole. Inoltre, l'applicazione di tali algoritmi compromette le prestazioni di algoritmi automatici. In questo lavoro proponiamo un sistema di trasformazione di immagini usando reti convoluzionali. Si dimostra che ottimizzando il modello tramite un approccio basato su GAN riusciamo a produrre immagini fotorealistiche, sulle quali gli algoritmi di visione continuano ad avere buone performance. Il nostro approccio integra un predittore di intensità della compressione per potere ricostruire qualsiasi tipo di immagine a prescindere dalla sua qualità.

1 Introduzione

Ogni giorno miliardi di immagini sono condivise sul web, e molte altre vengono memorizzate su dispositivi digitali. Per rendere pratici il salvataggio e la trasmissione di questi dati è necessario usare la compressione. Un tipico impiego di utilizzo è quello relativo alle immagini condivise in rete, infatti è necessario mantenere una bassa dimensione di questi file per ridurre al minimo la latenza di caricamento delle pagine web. Un altro frequente caso di utilizzo è quello relativo alle videocamere wireless, in particolare si fa riferimento a dispositivi mobili e indossabili, i quali hanno specifiche esigenze di ridurre al minimo il consumo di risorse. Infine gli streaming di video, per i quali c'è la necessità di limitare il più possibile l'utilizzo di banda per evitare la congestione della rete e ridurre i costi di trasmissione.

Purtroppo, quando viene utilizzato un algoritmo di compressione con perdita di dati, con alta probabilità compaiono artefatti indesiderati. Pertanto, è possibile osservare perdita di dettagli ad alta frequenza, strutture a forma di blocco e rumore. I principali svantaggi a seguito di questo fenomeno sono due. In primo luogo, le persone percepiscono le immagini compresse in modo meno gradevole. Inoltre, algoritmi di visione quali rilevatori di oggetti, segmentatori e captioner tendono ad avere scarse performance sulle immagini

comprese. Per questo motivo, si necessita di eliminare tali artefatti, invertendo il processo di compressione tramite una trasformazione non lineare dell'immagine.



Figura 1: Ricostruzione compressione: a sinistra un'immagine JPEG compressa a QF 20, a destra la ricostruzione dell'immagine con il nostro approccio.

1.1 Contributi

In questo lavoro proponiamo un metodo che impiega reti neurali residue interamente convoluzionali, che lavora in fase di post-processing delle immagini compresse. In questo modo è possibile applicare la tecnica per qualsiasi algoritmo di compressione di immagini, come JPEG, JPEG2000, WebP e anche compressione video, come H.264/AVC, VP9 e H.265/HEVC. Il nostro modello può essere ottimizzato in modo tradizionale, utilizzando una loss basata sulla similarità strutturale delle immagini (SSIM) e una basata sull'errore quadratico medio (MSE). Tuttavia, nonostante queste ottimizzazioni riescano a rimuovere la maggior parte degli artefatti di compressione, non riescono effettivamente a ricalcare la complessità strutturale del sistema percettivo umano. Per questo motivo, per apprendere migliori modelli ricostruttivi facciamo uso di Generative Adversarial Networks (GAN), le quali sono strumenti capaci di modellare complesse distribuzioni multimodali, pertanto in grado di generare immagini di migliore qualità. Il nostro approccio prevede inoltre un sistema capace di capire la qualità dell'immagine in modo automatico, quindi poter ricostruire a qualsiasi livello di compressione.

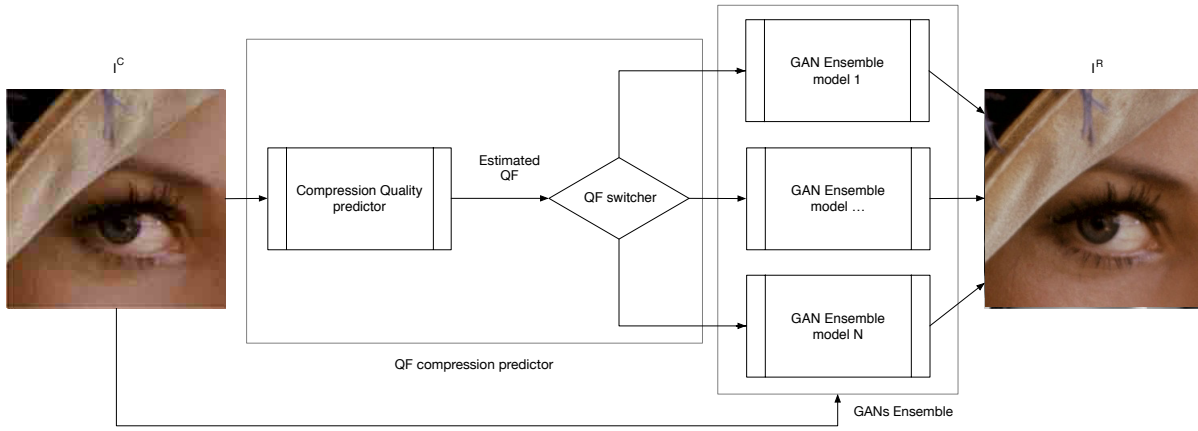


Figura 2: Schema del sistema: l'immagine compressa I^C è analizzata dal predittore di qualità. Il valore stimato è utilizzato per selezionare un appropriato modello GAN dall'ensemble per produrre l'immagine ricostruita I^R .

2 Metodo

La rimozione degli artefatti di compressione ha come obiettivo quello di ottenere una immagine di output I^R a partire da un'immagine compressa I^C , ottenuta tramite una trasformazione $I^C = A(I)$, dove I è l'immagine originale e A un qualsiasi algoritmo di compressione. Da questo punto di vista, il problema della ricostruzione di immagini compresse può essere interpretato come l'apprendimento della funzione inversa $G \approx A^{-1}$ che ricostruisce I a partire da I^C :

$$G(I^C) = I^R \approx I \quad (1)$$

La funzione si apprende mediante l'addestramento di una rete interamente convoluzionale, che può essere ottimizzata mediante supervisione diretta o combinata con un discriminatore, componendo in questo modo un sistema di apprendimento avversario.

2.1 Il generatore

Il generatore di immagini ricostruite è di fatto una rete convoluzionale basata su blocchi residui, i quali sono composti esclusivamente da convoluzioni e attivazioni non lineari. La peculiarità dei blocchi residui è che l'output consiste nella somma tra l'uscita dell'ultima convoluzione e l'input del blocco.

2.2 Il discriminatore

Il discriminatore è una rete neurale composta da una serie di layer convoluzionali e attivazioni non lineari. L'ultima uscita della rete ha dimensione unitaria alla quale è applicata una sigmoide come attivazione. A differenza delle tradizionali architetture dei discriminatori, il nostro modello è strutturato in modo da lavorare a livello di sub-patch. Ciò è motivato dal fatto che gli algoritmi di compressione tipicamente decompongono l'immagine a livello di patch, pertanto gli artefatti sono intrinsecamente legati ad esse. Prima di essere date in input al discriminatore, le immagini sono quindi suddivise in patch di piccole dimensioni.

2.3 Predizione di qualità

Poiché la qualità di un'immagine tipicamente non può essere conosciuta a priori, l'applicazione del nostro approccio in uno scenario realistico può essere problematica. Infatti, addestrare una GAN con immagini di qualsiasi qualità non è una soluzione percorribile, in quanto si osserva uno sgradevole fenomeno di *mode collapse* a causa delle compressioni molto aggressive. Per questo motivo, la nostra soluzione comprende due moduli che lavorano su piani differenti come mostrato in Fig. 2. La prima parte predice la qualità dell'immagine da ricostruire. Questo è possibile mediante l'addestramento di una rete apposita, che mediante una semplice regressione è in grado di stimare la qualità con precisione molto accurata. Il predittore è dunque impiegato per inviare l'immagine da ricostruire al modello GAN specializzato.

3 Conclusioni

In questo documento è stato descritto l'approccio per la rimozione di artefatti da immagini compresse, ottenuto mediante l'apprendimento di una funzione di trasformazione dell'immagine mediante una rete neurale. Il sistema basato su GAN è in grado di produrre immagini con dettagli molto fini che sono importanti non solo da un punto di vista di apprezzabilità da parte dell'uomo, ma anche per gli algoritmi automatici di visione. Considerando che i parametri di compressione non sono noti a priori, il metodo descritto è in grado di predire la qualità dell'immagine e scegliere un modello specializzato per una ricostruzione ottimale.

Riferimenti bibliografici

- [Galteri *et al.*, 2017] Leonardo Galteri, Lorenzo Seidenari, Marco Bertini, e Alberto Del Bimbo. Deep generative adversarial compression artifact removal. In *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.
- [Galteri *et al.*, 2019] Leonardo Galteri, Lorenzo Seidenari, Marco Bertini, e Alberto Del Bimbo. Deep universal generative adversarial compression artifact removal. *IEEE Transactions on Multimedia*, 2019.