

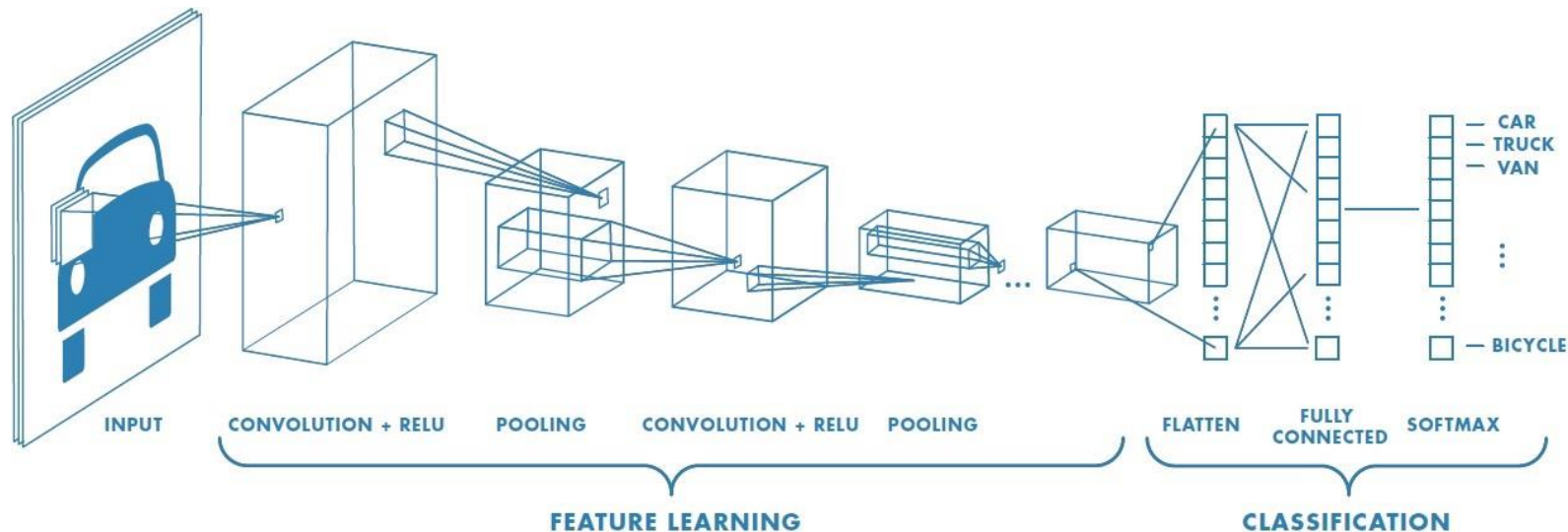


SVILUPPO, ANALISI E IMPLEMENTAZIONE DI UN ALGORITMO OTTIMIZZATO PER LE RETI DI CONVOLUZIONE


Realizzato da:
Al Sadi Amir
Saccone Nicolò

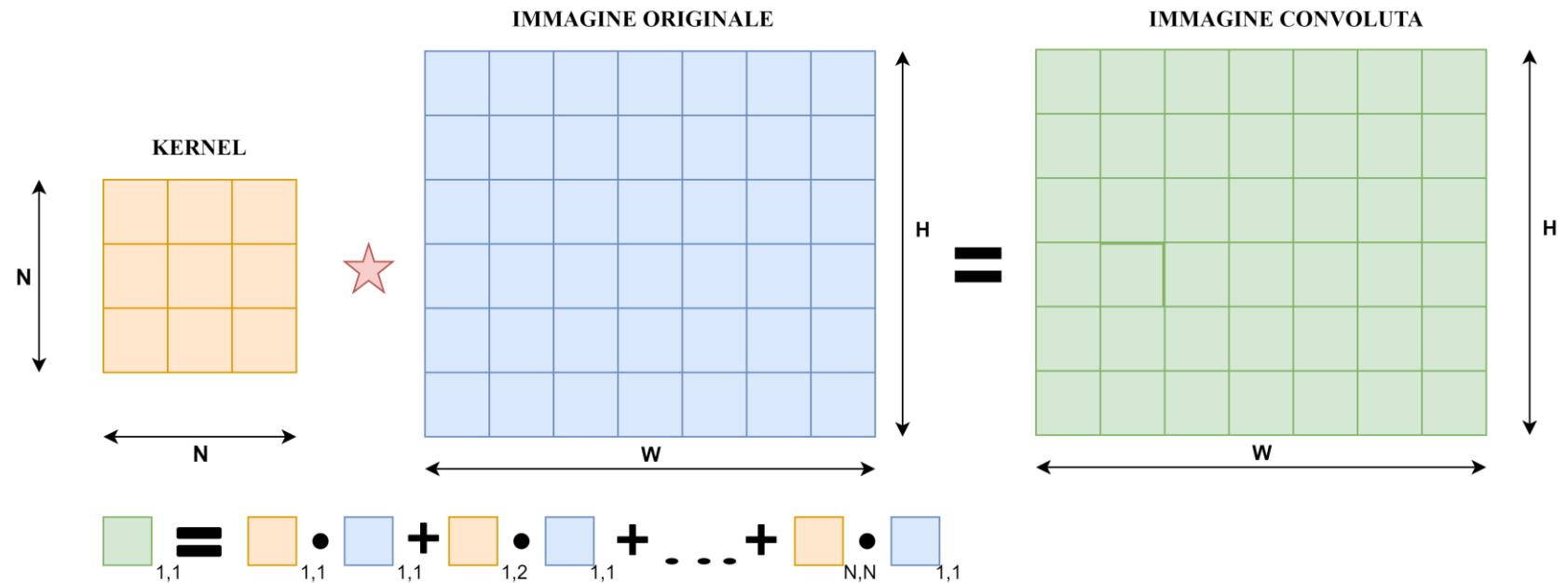
CNN: CONVOLUTIONAL NEURAL NETWORK

- Con CNN si intende una categoria di reti neurali largamente utilizzata in ambito di applicazioni di visione artificiale in grado di elaborare segnali di diverso tipo come immagini o suoni.
- Una rete di convoluzione neurale è solitamente composta da qualche migliaia di livelli detti layer di convoluzione aventi ognuno specifiche caratteristiche e finalità.

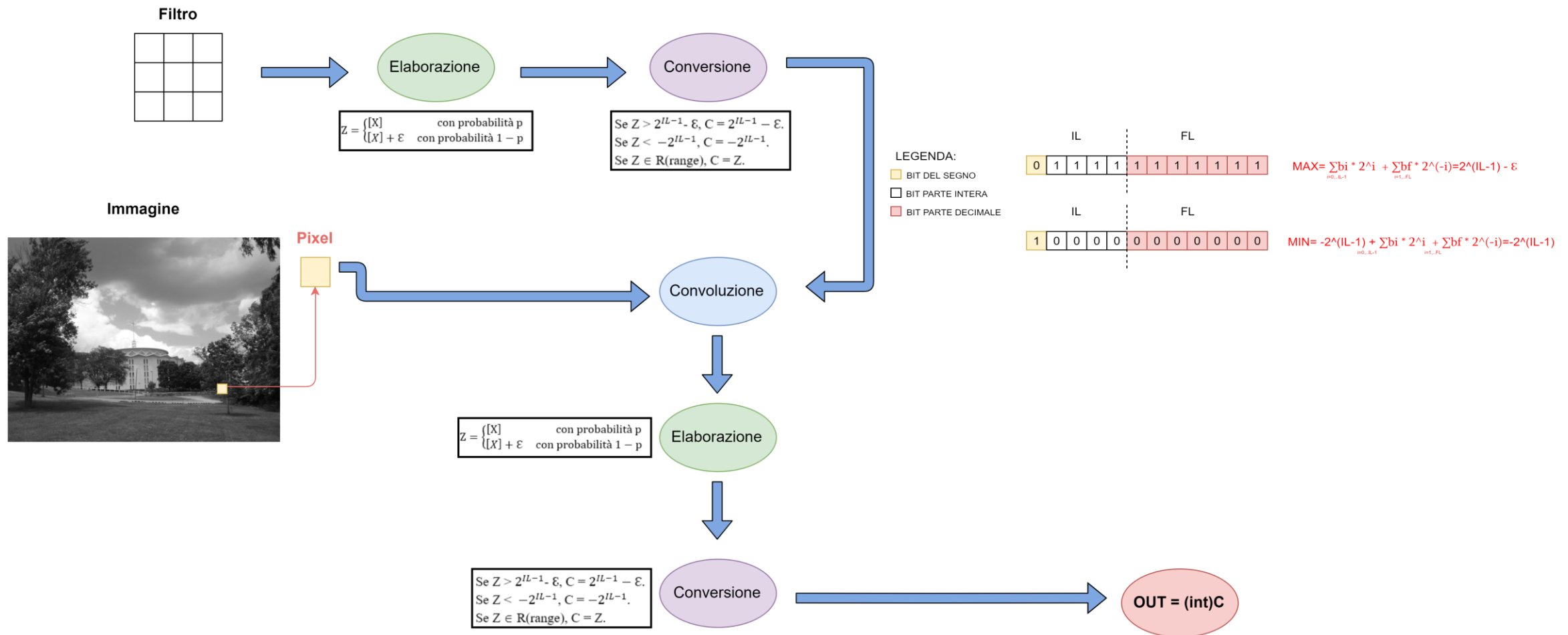


LAYER DI CONVOLUZIONE

- Un layer di convoluzione è uno dei tanti elementi che compongono una CNN. Comunemente esso ha come ingressi:
 - Un'immagine sottoforma di matrice di pixel.
 - Una matrice di convoluzione detta kernel, i cui valori determinano l'effetto in uscita.
- L'operazione , chiamata convoluzione, determina una nuova immagine in uscita di dimensione pari all'ingresso.



ALGORITMO DI MAPPING SU INTERI



CASO DI STUDIO

- Immagini in input di dimensione 640x480 (formato compatibile con VGA) con pixel a 8-bit grayscale.
- Kernel di convoluzione di dimensione 3x3 e 7x7 e separabili , per poter analizzare i risultati dell'algoritmo in entrambi le versioni: separabile e standard.
- Valori dei kernel rappresentati da float a 32 bit, ridotti tramite l'algoritmo stocastico.
- Report delle risorse utilizzate da un layer tramite il tool Vivado HLS e analisi della qualità delle immagini in uscita allo stesso tramite gli indici MSE, RMSE e SSIM.

ANALISI RISORSE UTILIZZATE CON KERNEL 3X3

- Utilizzo di un kernel gaussiano separabile.
- Grosso guadagno di memoria rispetto alla convoluzione senza algoritmo stocastico.
- Trend in calo non soddisfacente del consumo di memoria al variare del numero di bit allocati per i dati.

Motivi:

- Vivado HLS.
- Cast interni al codice.
- Matrice di appoggio per la convoluzione.
- Valori assegnati a IL e FL.

<div> <div>LEGENDA</div> <div> <div>CONVOLUZIONE TRADIZIONALE</div> <div>CONVOLUZIONE CON KERNEL SEPARABILE</div> <div>RIGA SENZA OTTIMIZZAZIONE (CONV. con FLOAT)</div> </div> <div>CASO KERNEL 3x3</div> </div>										
RISORSE UTILIZZATE N° BIT	LATENCY (Clock cycles)	LATENCY (Clock cycles)	CLOCK ESTIMATED (ns)	CLOCK ESTIMATED (ns)	BRAM (18k)	BRAM (18k)	FF(106400 DISPONIBILI)	FF(106400 DISPONIBILI)	LUT(53200 DISPONIBILI)	LUT(53200 DISPONIBILI)
10	308371	308372	8.585	8.585	2	2	3209	3255	6560	6446
12	308371	308372	8.609	8.585	2	2	3209	3262	6562	6458
14	308371	308372	8.638	8.585	2	2	3109	3280	6564	6494
16	308371	308372	8.638	8.585	2	2	3198	3298	6564	6511
24	308372	308372	8.585	8.585	2	2	3225	3507	6568	6559
32	308371	308371	8.585	8.585	2	2	3208	3059	6677	6407
32	308435	308412	8.666	8.666	2	2	11167	7564	18756	13453

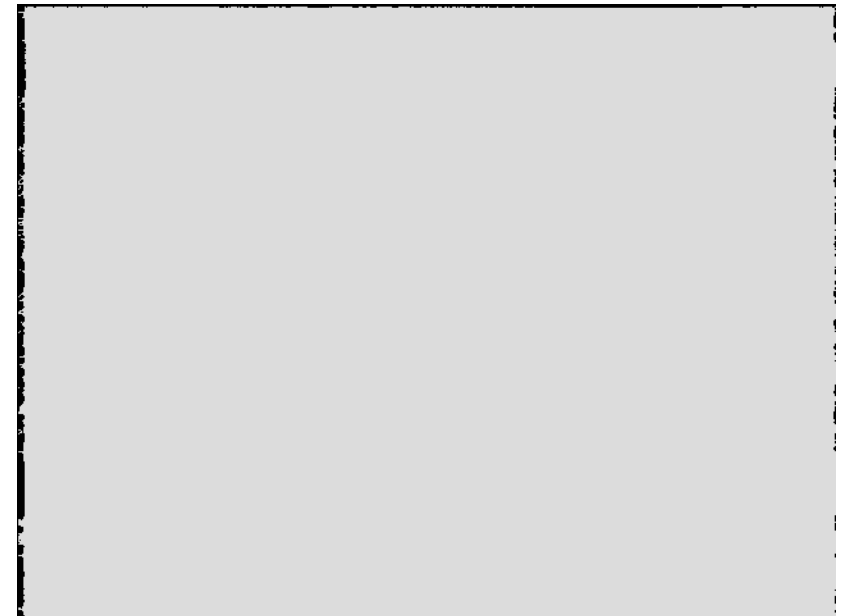
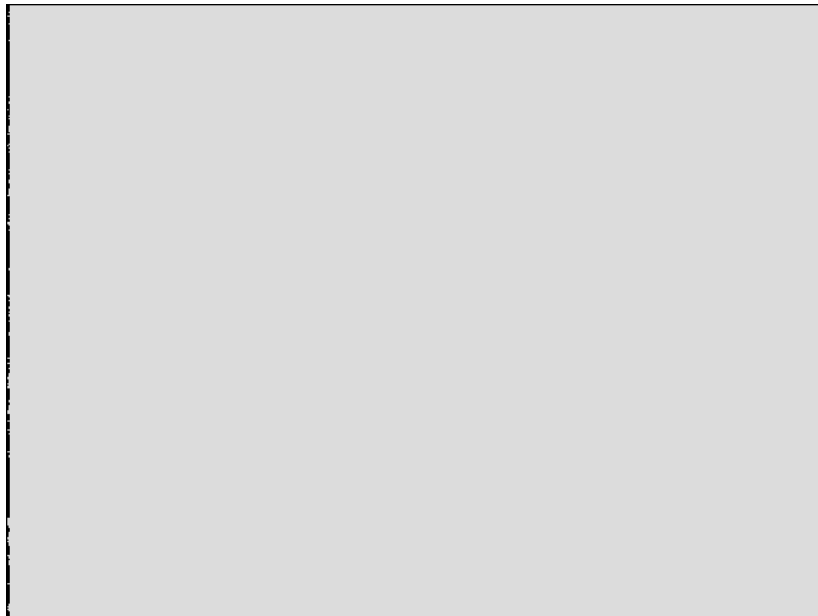
ANALISI RISORSE UTILIZZATE CON KERNEL 7X7

- Utilizzo di un kernel gaussiano separabile.
- I risultati riconfermano le supposizioni nel caso 3x3 nel confronto fra l'algoritmo classico e stocastico.
- Trend leggermente più marcato rispetto al caso 3x3.
- A un maggior numero di moltiplicazioni corrisponde un consumo più elevato di memoria.

<div> <div>LEGENDA</div> <div> <div>CONVOLUZIONE TRADIZIONALE</div> <div>CONVOLUZIONE CON KERNEL SEPARABILE</div> <div>RIGA SENZA OTTIMIZZAZIONE (CONV. con FLOAT)</div> </div> <div>CASO KERNEL 7x7</div> </div>										
RISORSE UTILIZZATE N° BIT	LATENCY (Clock cycles)	LATENCY (Clock cycles)	CLOCK ESTIMATED (ns)	CLOCK ESTIMATED (ns)	BRAM (18k)	BRAM (18k)	FF(106400 DISPONIBILI)	FF(106400 DISPONIBILI)	LUT(53200 DISPONIBILI)	LUT(53200 DISPONIBILI)
10	310620	310621	8.655	8.585	6	6	3688	3255	7254	6560
12	310620	310620	8.655	8.609	6	6	3690	3262	7256	6562
14	310620	310620	8.655	8.638	6	6	3692	3280	7258	6564
16	310620	310620	8.698	8.638	6	6	3684	3298	7243	6564
24	310620	310621	8.698	8.585	6	6	3684	3507	7243	6568
32	308371	310619	8.585	8.742	6	6	3746	3203	8073	6695
32	311252	310768	8.666	8.666	6	6	68349	19088	119461	33202

ANALISI QUALITATIVA DELLE IMMAGINI

- Confronto puntuale dei valori dei pixel dell'immagine in uscita dalla convoluzione a piena precisione realizzata con il tool Vivado HLS con quelli dell'immagine convoluta tramite Open CV.
- Colorazione in grigio dei pixel che hanno differenza in valore assoluto di massimo 2 unità ed in nero gli altri.
- In entrambi i casi 3x3(sinistra) e 7x7(destra) i risultati presentano delle disparità sui bordi.



ANALISI QUALITATIVA CON KERNEL 3X3

- Utilizzo di IL e FL differenti fra caso tradizionale e separabile a causa delle diverse approssimazioni conseguenti ai due metodi, per ottenere i migliori risultati in uscita.
- Somiglianza e distanza assoluta delle immagini soddisfacenti dai 32 ai 14 bit.
- Differenza di qualità maggiormente delineata a partire da 12 bit di approssimazione, a favore dell'algoritmo che non sfrutta la separabilità del kernel.

LEGENDA

CONVOLUZIONE TRADIZIONALE

CONVOLUZIONE CON KERNEL SEPARABILE

CASO KERNEL
3x3

QUALITÀ IMMAGINE		MSE	RMSE	SSIM	QUALITÀ IMMAGINE		MSE	RMSE	SSIM
N° BIT	IL				FL	N° BIT			
16	16	2.79	1.67	1.00	28	4	10.11	3.18	1.00
14	10	2.79	1.67	1.00	20	4	10.11	3.18	1.00
10	6	2.79	1.67	1.00	12	4	10.11	3.18	1.00
10	4	2.79	1.67	1.00	10	4	10.11	3.18	1.00
10	2	958.99	30.97	0.92	8	4	1356.57	36.83	0.89
8	2	1714.08	41.40	0.86	6	4	12793.68	113.11	0.27

ANALISI QUALITATIVA CON KERNEL 3X3



Classico



Separabile

ANALISI QUALITATIVA CON KERNEL 7X7

- Riconfermate le osservazioni del caso 3x3.
- A un aumento dei prodotti corrisponde un peggioramento della qualità dell'immagine.
- L'effetto di peggioramento si riflette maggiormente nell'algoritmo separabile.

<div> <div>LEGENDA</div> <div> <div>CONVOLUZIONE TRADIZIONALE</div> <div>CONVOLUZIONE CON KERNEL SEPARABILE</div> </div> <div>CASO KERNEL 7x7</div> </div>									
QUALITÀ IMMAGINE		MSE	RMSE	SSIM	QUALITÀ IMMAGINE		MSE	RMSE	SSIM
N° BIT	IL FL				N° BIT	IL FL			
16	16	5.29	2.3	1.00	22	10	5.11	2.26	1.00
10	14	5.29	2.3	1.00	14	10	5.11	2.26	1.00
10	6	2114.68	45.99	0.94	8	8	1328.66	36.45	0.90
10	4	3043.70	55.17	0.89	8	6	1343.64	36.66	0.88
10	2	3731.88	61.09	0.72	8	4	1418.83	37.67	0.88
8	2	1418.83	37.67	0.85	6	4	3227.36	56.81	0.41

ANALISI QUALITATIVA CON KERNEL 7X7



Classico



Separabile

IMPLEMENTAZIONE SU FPGA

- Si è effettuato un deploy dell'algoritmo su una Zynq-7000 ZedBoard per verificare il funzionamento dello stesso in un caso reale e valutarne la qualità complessiva ottenuta rispetto al caso non ottimizzato.
- L'esperimento è stato effettuato tramite l'utilizzo del tool Vivado sulla base di un progetto preesistente(<https://github.com/smatt-github/SmartCamera>) che realizza diverse funzionalità fra le quali lo stream su VGA delle immagini filtrate acquisite dalla telecamera OV7670.
- Riadattamento del modulo "Filter Convolution".
- I risultati visivi riconfermano le supposizioni fatte utilizzando il tool OpenCV su un PC.

CONCLUSIONI

- L'utilizzo dell'algoritmo proposto consente un grosso risparmio di risorse allocate rispetto ai suoi equivalenti a massima precisione.
- La validità dell'algoritmo si estende soprattutto in ottica di training, in quanto a un grosso guadagno di memoria è associata una ottima qualità dell'immagine.
- I possibili sviluppi futuri sono:
 - Ampliamento dei test e aggiustamenti del codice HLS.
 - Utilizzo di diversi filtri e confronto fra di essi, per stabilire l'influenza di un determinato filtro sulla rete.
 - Realizzazione di un design implementativo progettato utilizzando le specifiche indicate nel progetto iniziale.