(Hw 5)

(Ex 1.2)

A gray scale image is 2D $\Rightarrow$ Just a matrix

Using a patch-wise DCT transform with 4x4 patches we have for each pixel $u(x,y)$ of the 4x4 patch, [illegible]

$$U(x,y) = \sum_{0 \leq i,j \leq 3} 4\alpha_i \alpha_j \cos\left(\pi\left(i+\tfrac{1}{2}\right)\tfrac{x}{4}\right) \cos\left(\pi\left(j+\tfrac{1}{2}\right)\tfrac{y}{4}\right) u(i,j)$$

$=$ Convolution between u and

$$K_{x,y}(i,j) = 4\alpha_i\alpha_j \cos\left(\pi\left(i+\tfrac{1}{2}\right)\tfrac{x}{4}\right) \cos\left(\pi\left(j+\tfrac{1}{2}\right)\tfrac{y}{4}\right)$$

(we adopt the convolution def as seen in Convolutional layer

i.e $(x * y)(i,j) = \sum_{k,l} x(k,l)\, y(k,l)$)

Thus, we can implement the patch wise DCT transform using 16 kernels $K_{x,y}(i,j)$ for $0 \leq x,y \leq 3$,

$$\frac{\partial f_1}{\partial x_2}(x;\theta_1) \frac{\partial f_2}{\partial x_1}(f_1(x;\theta_1),\theta_2)\left[1 + \frac{\partial f_3}{\partial x_1}(y;\theta_3)\right] =$$

(Ex 1.4)

- $\dfrac{\partial F}{\partial \theta_3}(x) = \dfrac{\partial f_3}{\partial x_2}(y;\theta_3)$
- $\dfrac{\partial F}{\partial \theta_2}(x) = \dfrac{\partial y}{\partial \theta_2} \dfrac{\partial f_3}{\partial x_1}(y;\theta_3)$

$$= \frac{\partial f_2}{\partial x_2}(f_1(x;\theta_1),\theta_2)$$

$$\Rightarrow \frac{\partial F}{\partial \theta_2}(x) = \frac{\partial f_2}{\partial x_2}(f_1(x;\theta_1),\theta_2) \frac{\partial f_3}{\partial x_1}(y;\theta_3)$$

- $\dfrac{\partial F}{\partial \theta_1}(x) = \dfrac{\partial y}{\partial \theta_1} \dfrac{\partial f_3}{\partial x_1}(y;\theta_3)$

$$= \frac{\partial f_1}{\partial x_2}(x,\theta_1) \frac{\partial f_2}{\partial x_1}(f_1(x;\theta_1),\theta_2) \frac{\partial f_3}{\partial x_1}(y;\theta_3)$$

Where we adopt the notation $\dfrac{\partial f}{\partial x_i}(.,.)$ for the partial derivative w.r.t the i-th component of f.

- $\dfrac{\partial G}{\partial \theta_3}(x) = \dfrac{\partial F}{\partial \theta_3}(x) = \dfrac{\partial f_3}{\partial x_2}(y;\theta_3)$
- $\dfrac{\partial G}{\partial \theta_2}(x) = \dfrac{\partial y}{\partial \theta_2} + \dfrac{\partial F(x)}{\partial \theta_2}$

$$= \frac{\partial f_2}{\partial x_2}(f_1(x;\theta_1),\theta_2) \left(1 + \frac{\partial f_3}{\partial x_1}(y;\theta_2)\right)$$

- $\dfrac{\partial G}{\partial \theta_1}(x) = \dfrac{\partial y}{\partial \theta_1}\left(1 + \dfrac{\partial f_3}{\partial x_1}(y,\theta_2)\right)$

For $\dfrac{\partial G}{\partial \theta_1}$, we see the presence of $\dfrac{\partial y}{\partial \theta_1}$ which doesn't depend on f_3 $\Rightarrow$ we offer an additional path for the gradient flow in case f_3 has vanishing gradients!