Supplement: Discrete Deep Feature Extraction: A Theory and New Architectures

Thomas Wiatowski[1] WITHOMAS@NARI.EE.ETHZ.CH
Michael Tschannen[1] MICHAELT@NARI.EE.ETHZ.CH
Aleksandar Stanić[1] ASTANIC@STUDENT.ETHZ.CH
Philipp Grohs[2] PHILIPP.GROHS@UNIVIE.AC.AT
Helmut Bölcskei[1] BOELCSKEI@NARI.EE.ETHZ.CH
[1]Dept. IT & EE, ETH Zurich, Switzerland
[2]Dept. Math., University of Vienna, Austria

A. Appendix: Additional numerical results

A.1. Handwritten digit classification

For the handwritten digit classification experiment described in Section 6.1, Table 1 shows the classification error for Daubechies wavelets with 2 vanishing moments (DB2).

	DB2			
	abs	ReLU	tanh	LogSig
n.p.	0.54	0.51	1.29	1.40
sub.	0.60	0.58	1.16	1.34
max.	0.57	0.57	0.75	0.67
avg.	0.52	0.61	1.16	1.27

Table 1. Classification errors in percent for handwritten digit classification using DB2 wavelet filters, different non-linearities, and different pooling operators (sub.: sub-sampling; max.: max-pooling; avg.: average-pooling; n.p.: no pooling).

A.2. Feature importance evaluation

For the feature importance experiment described in Section 6.2, Figure 1 shows the cumulative feature importance (per triplet of layer index, wavelet scale, and direction, averaged over all trees in the respective RF) in facial landmark detection (right eye and mouth).

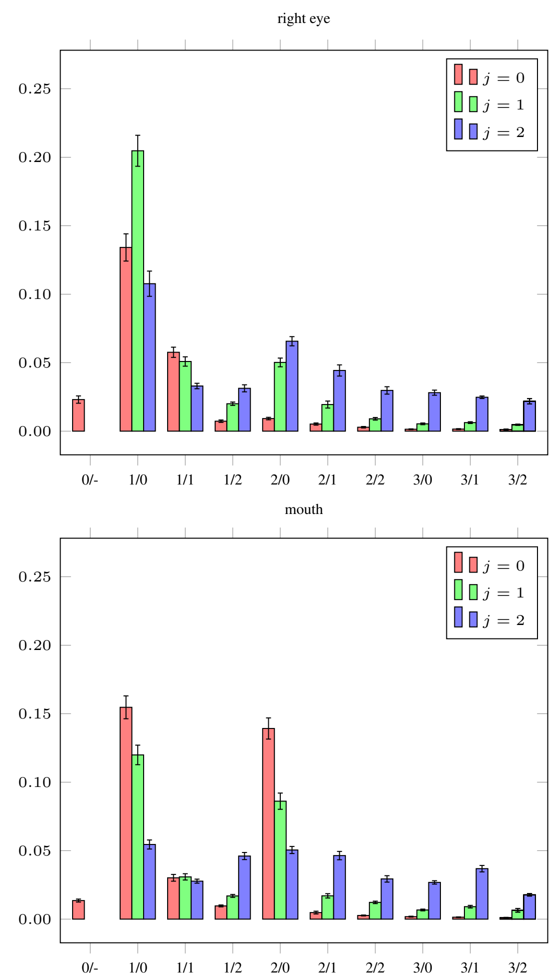

Figure 1. Average cumulative feature importance and standard error for facial landmark detection. The labels on the horizontal axis indicate layer index d/wavelet direction (0: horizontal, 1: vertical, 2: diagonal).

B. Appendix: Lipschitz continuity of pooling operators

We verify the Lipschitz property

$$\|P(f) - P(h)\|_2 \leq R\|f - h\|_2, \quad \forall f, h \in H_N,$$

for the pooling operators in Section 2.3.1.

Sub-sampling: Pooling by sub-sampling is defined as

$$P : H_N \rightarrow H_{N/S}, \quad P(f)[n] = f[Sn], \quad n \in I_{N/S},$$

where $N/S \in \mathbb{N}$. Lipschitz continuity with $R = 1$ follows from

$$\|P(f) - P(h)\|_2^2 = \sum_{n \in I_{N/S}} |f[Sn] - h[Sn]|^2$$
$$\leq \sum_{n \in I_N} |f[n] - h[n]|^2 = \|f - h\|_2^2, \quad \forall f, h \in H_N.$$

Averaging: Pooling by averaging is defined as

$$P : H_N \to H_{N/S}, \quad P(f)[n] = \sum_{k=Sn}^{Sn+S-1} \alpha_{k-Sn} f[k],$$

for $n \in I_{N/S}$, where $N/S \in \mathbb{N}$. We start by setting $\alpha' := \max_{k\in\{0,\dots,S-1\}} |\alpha_k|$. Then,

$$\|P(f) - P(h)\|_2^2$$
$$= \sum_{n \in I_{N/S}} \Big| \sum_{k=Sn}^{Sn+S-1} \alpha_{k-Sn}(f[k] - h[k]) \Big|^2$$
$$\leq \sum_{n \in I_{N/S}} \Big| \sum_{k=Sn}^{Sn+S-1} \alpha' |f[k] - h[k]| \Big|^2$$
$$\leq \alpha'^2 S \sum_{n \in I_{N/S}} \sum_{k=Sn}^{Sn+S-1} \Big| f[k] - h[k] \Big|^2 \tag{B.1}$$
$$= \alpha'^2 S \sum_{n \in I_N} \Big| f[k] - h[k] \Big|^2 = \alpha'^2 S \|f - h\|_2^2,$$

where we used $\sum_{k\in I_S} |f[k]-h[k]| \leq S^{1/2}\|f-h\|_2$, $f, h \in H_S$, to get (B.1), see, e.g., (Golub & Van Loan, 2013).

Maximization: Pooling by maximization is defined as

$$P : H_N \to H_{N/S}, \quad P(f)[n] = \max_{k\in\{Sn,\dots,Sn+S-1\}} |f[k]|,$$

for $n \in I_{N/S}$, where $N/S \in \mathbb{N}$. We have

$$\|P(f) - P(h)\|_2^2$$
$$= \sum_{n \in I_{N/S}} \Big| \max_{k\in\{Sn,\dots,Sn+S-1\}} |f[k]|$$
$$- \max_{k\in\{Sn,\dots,Sn+S-1\}} |h[k]| \Big|^2$$
$$\leq \sum_{n \in I_{N/S}} \max_{k\in\{Sn,\dots,Sn+S-1\}} \Big| f[k] - h[k] \Big|^2 \tag{B.2}$$
$$\leq \sum_{n \in I_{N/S}} \sum_{k=0}^{S-1} |f[Sn+k] - h[Sn+k]|^2 \tag{B.3}$$
$$= \|f - h\|_2^2,$$

where we employed the reverse triangle inequality $\big|\|f\|_\infty - \|h\|_\infty\big| \leq \|f - h\|_\infty$, $f, h \in H_S$, to get (B.2), and in (B.3) we used $\|f\|_\infty \leq \|f\|_2$, $f \in H_S$, see, e.g., (Golub & Van Loan, 2013).

C. Appendix: Proof of Theorem 1

We start by proving i). The key idea of the proof is—similarly to the proof of Proposition 4 in (Wiatowski & Bölcskei, 2015)—to employ telescoping series arguments. For ease of notation, we let $f_q := U[q]f$ and $h_q := U[q]h$, for $f, h \in H_{N_1}$, $q \in \Lambda_1^d$. With (9) we have

$$|||\Phi_\Omega(f) - \Phi_\Omega(h)|||^2 = \sum_{d=0}^{D-1} \underbrace{\sum_{q\in\Lambda_1^d} ||(f_q - h_q) * \chi_d||_2^2}_{=:a_d}.$$

The key step is then to show that a_d can be upper-bounded according to

$$a_d \leq b_d - b_{d+1}, \qquad d = 0, \dots, D-1, \tag{C.1}$$

with $b_d := \sum_{q\in\Lambda_1^d} \|f_q - h_q\|_2^2$, for $d = 0, \dots, D$, and to note that

$$\sum_{d=0}^{D-1} a_d \leq \sum_{d=0}^{D-1} (b_d - b_{d+1}) = b_0 - \underbrace{b_D}_{\geq 0} \leq b_0$$
$$= \sum_{q\in\Lambda_1^0} \|f_q - h_q\|_2^2 = \|f - h\|_2^2,$$

which then yields (8). Writing out (C.1), it follows that we need to establish

$$\sum_{q\in\Lambda_1^d} \|(f_q - h_q) * \chi_d\|_2^2 \leq \sum_{q\in\Lambda_1^d} ||f_q - h_q||_2^2$$
$$- \sum_{q\in\Lambda_1^{d+1}} \|f_q - h_q\|_2^2, \qquad d = 0, \dots, D-1. \tag{C.2}$$

We start by examining the second sum on the right-hand side (RHS) in (C.2). Every path

$$\tilde{q} \in \Lambda_1^{d+1} = \underbrace{\Lambda_1 \times \cdots \times \Lambda_d}_{=\Lambda_1^d} \times \Lambda_{d+1}$$

of length $d+1$ can be decomposed into a path $q \in \Lambda_1^d$ of length d and an index $\lambda_{d+1} \in \Lambda_{d+1}$ according to $\tilde{q} = (q, \lambda_{d+1})$. Thanks to (5) we have $U[\tilde{q}] = U[(q, \lambda_{d+1})] = U_{d+1}[\lambda_{d+1}]U[q]$, which yields

$$\sum_{\tilde{q}\in\Lambda_1^{d+1}} \|f_{\tilde{q}} - h_{\tilde{q}}\|_2^2 = \sum_{q\in\Lambda_1^d} \sum_{\lambda_{d+1}\in\Lambda_{d+1}} \|U_{d+1}[\lambda_{d+1}]f_q$$
$$- U_{d+1}[\lambda_{d+1}]h_q\|_2^2. \tag{C.3}$$

Substituting (C.3) into (C.2) and rearranging terms, we obtain

$$\sum_{q\in\Lambda_1^d}\Big(\|(f_q-h_q)*\chi_d\|_2^2 \tag{C.4}$$

$$+\sum_{\lambda_{d+1}\in\Lambda_{d+1}}\|U_{d+1}[\lambda_{d+1}]f_q-U_{d+1}[\lambda_{d+1}]h_q\|_2^2\Big) \tag{C.5}$$

$$\le \sum_{q\in\Lambda_1^d}\|f_q-h_q\|_2^2, \qquad d=0,\dots,D-1. \tag{C.6}$$

We next note that the sum over the index set Λ_{d+1} inside the brackets in (C.4)-(C.5) satisfies

$$\sum_{\lambda_{d+1}\in\Lambda_{d+1}}\|U_{d+1}[\lambda_{d+1}]f_q-U_{d+1}[\lambda_{d+1}]h_q\|_2^2$$

$$=\sum_{\lambda_{d+1}\in\Lambda_{d+1}}\|P_{d+1}\big(\rho_{d+1}(f_q*g_{\lambda_{d+1}})\big)$$

$$-P_{d+1}\big(\rho_{d+1}(h_q*g_{\lambda_{d+1}})\big)\|_2^2$$

$$\le R_{d+1}^2\sum_{\lambda_{d+1}\in\Lambda_{d+1}}\|\rho_{d+1}(f_q*g_{\lambda_{d+1}}) \tag{C.7}$$

$$-\rho_{d+1}(h_q*g_{\lambda_{d+1}})\|_2^2 \tag{C.8}$$

$$\le R_{d+1}^2L_{d+1}^2\sum_{\lambda_{d+1}\in\Lambda_{d+1}}\|(f_q-h_q)*g_{\lambda_{d+1}}\|_2^2, \tag{C.9}$$

where we employed the Lipschitz continuity of P_{d+1} in (C.7)-(C.8) and the Lipschitz continuity of ρ_{d+1} in (C.9). Substituting the sum over the index set Λ_{d+1} inside the brackets in (C.4)-(C.5) by the upper bound (C.9) yields

$$\sum_{q\in\Lambda_1^d}\Big(\|(f_q-h_q)*\chi_d\|_2^2$$

$$+\sum_{\lambda_{d+1}\in\Lambda_{d+1}}\|U_{d+1}[\lambda_{d+1}]f_q-U_{d+1}[\lambda_{d+1}]h_q\|_2^2\Big)$$

$$\le\sum_{q\in\Lambda_1^d}\max\{1,R_{d+1}^2L_{d+1}^2\}\Big(\|(f_q-h_q)*\chi_d\|_2^2 \tag{C.10}$$

$$+\sum_{\lambda_{d+1}\in\Lambda_{d+1}}\|(f_q-h_q)*g_{\lambda_{d+1}}\|_2^2\Big), \tag{C.11}$$

for $d=0,\dots,D-1$. As $\{g_{\lambda_{d+1}}\}_{\lambda_{d+1}\in\Lambda_{d+1}}\cup\{\chi_d\}$ are atoms of the convolutional set Ψ_{d+1}, and $f_q,h_q\in H_{N_{d+1}}$, we have

$$\|(f_q-h_q)*\chi_d\|_2^2+\sum_{\lambda_{d+1}\in\Lambda_{d+1}}\|(f_q-h_q)*g_{\lambda_{d+1}}\|_2^2$$

$$\le B_{d+1}\|f_q-h_q\|_2^2,$$

which, when used in (C.10)-(C.11) yields

$$\sum_{q\in\Lambda_1^d}\Big(\|(f_q-h_q)*\chi_d\|_2^2$$

$$+\sum_{\lambda_{d+1}\in\Lambda_{d+1}}\|U_{d+1}[\lambda_{d+1}]f_q-U_{d+1}[\lambda_{d+1}]h_q\|_2^2\Big)$$

$$\le\sum_{q\in\Lambda_1^d}\max\{B_{d+1},B_{d+1}R_{d+1}^2L_{d+1}^2\}\|f_q-h_q\|_2^2, \tag{C.12}$$

for $d=0,\dots,D-1$. Finally, invoking (7) in (C.12) we get (C.4)-(C.6) and hence (C.1). This completes the proof of i).

We continue with ii). The key step in establishing (10) is to show that for $\rho_d(0)=0$ and $P_d(0)=0$, for $d\in\{1,\dots,D-1\}$, the feature extractor Φ_Ω satisfies $\Phi_\Omega(0)=0$, and to employ (8) with $h=0$ which yields

$$|||\Phi(f)|||\le\|f\|,$$

for $f\in H_{N_1}$. It remains to prove that $\Phi_\Omega(h)=0$ for $h=0$. For $h=0$, the operator U_d, $d\in\{1,2,\dots,D\}$, defined in (4) satisfies

$$(U_d[\lambda_d]h)=\underbrace{P_d\big(\underbrace{\rho_d(\underbrace{h*g_{\lambda_d}}_{=0})}_{=0}\big)}_{=0},$$

for $\lambda_d\in\Lambda_d$, by assumption. With the definition of $U[q]$ in (5) this then yields $(U[q]h)=0$ for $h=0$ and all $q\in\Lambda_1^d$. $\Phi_\Omega(0)=0$ finally follows from

$$\Phi_\Omega(h)=\bigcup_{d=0}^{D-1}\big\{\underbrace{(U[q]h)*\chi_d}_{=0}\big\}_{q\in\Lambda_1^d}=0. \tag{C.13}$$

We proceed to iii). The proof of the deformation sensitivity bound (12) is based on two key ingredients. The first one is the Lipschitz continuity result stated in (8). The second ingredient, stated in Proposition D.1 in Appendix D, is an upper bound on the deformation error $\|f-F_\tau f\|_2$ given by

$$\|f-F_\tau f\|_2\le 4KN_1^{1/2}\|\tau\|_\infty^{1/2}, \tag{C.14}$$

where $f\in C_{\text{CART}}^{N_1,K}$. We now show how (8) and (C.14) can be combined to establish (12). To this end, we first apply (8) with $h:=(F_\tau f)$ to get

$$|||\Phi_\Omega(f)-\Phi_\Omega(F_\tau f)|||\le\|f-F_\tau f\|_2, \tag{C.15}$$

for $f\in C_{\text{CART}}^{N_1,K}\subseteq H_{N_1}$, $N_1\in\mathbb{N}$, and $K>0$, and then replace the RHS of (C.15) by the RHS of (C.14). This completes the proof of iii).

D. Appendix: Proposition D.1

Proposition D.1. *For every $N \in \mathbb{N}$, every $K > 0$, and every $\tau : \mathbb{R} \to [-1, 1]$, we have*

$$\|f - F_\tau f\|_2 \le 4KN^{1/2}\|\tau\|_\infty^{1/2}, \tag{D.1}$$

for all $f \in \mathcal{C}^{N,K}_{\mathrm{CART}}$.

Remark D.1. *As already mentioned at the end of Section 4, excluding the interval boundary points a, b in the definition of sampled cartoon functions $\mathcal{C}^{N,K}_{\mathrm{CART}}$ (see Definition 4) is necessary for technical reasons. Specifically, without imposing this exclusion, we can not expect to get deformation sensitivity results of the form (D.1). This can be seen as follows. Let us assume that we seek a bound of the form $\|f - F_\tau f\|_2 \le C_{N,K}\|\tau\|^\alpha_\infty$, for some $C_{N,K} > 0$ and some $\alpha > 0$, that applies to all $f[n] = c(n/N)$, $n \in I_N$, with $c \in \mathcal{C}^K_{\mathrm{CART}}$. Take $\tau(x) = 1/N$, in which case the deformation $(F_\tau f)[n] = c(n/N - 1/N)$ amounts to a simple translation by $1/N$ and $\|\tau\|_\infty = 1/N \le 1$. Let $c(x) = \mathbb{1}_{[0,2/N]}(x)$. Then $c \in \mathcal{C}^K_{\mathrm{CART}}$ for $K = 1$ and $\|f - F_\tau f\|_2 = \sqrt{2}$, which obviously does not decay with $\|\tau\|^\alpha_\infty = N^{-\alpha}$ for some $\alpha > 0$. We note that this phenomenon occurs only in the discrete case.*

Proof. The proof of (D.1) is based on judiciously combining deformation sensitivity bounds for the sampled components $c_1(n/N), c_2(n/N)$, $n \in I_N$, in $(c_1 + \mathbb{1}_{[a,b]}c_2) \in \mathcal{C}^K_{\mathrm{CART}}$, and the sampled indicator function $\mathbb{1}_{[a,b]}(n/N)$, $n \in I_N$. The first bound, stated in Lemma D.1 below, reads

$$\|f - F_\tau f\|_2 \le CN^{1/2}\|\tau\|_\infty, \tag{D.2}$$

and applies to discrete-time signals $f[n] = f(n/N)$, $n \in I_N$, with $f : \mathbb{R} \to \mathbb{C}$ satisfying the Lipschitz property with Lipschitz constant C. The second bound we need, stated in Lemma D.2 below, is given by

$$\|\mathbb{1}^N_{[a,b]} - F_\tau \mathbb{1}^N_{[a,b]}\|_2 \le 2N^{1/2}\|\tau\|^{1/2}_\infty, \tag{D.3}$$

and applies to sampled indicator functions $\mathbb{1}^N_{[a,b]}[n] := \mathbb{1}_{[a,b]}(n/N)$, $n \in I_N$, with $a, b \notin \{0, \frac{1}{N}, \dots, \frac{N-1}{N}\}$. We now show how (D.2) and (D.3) can be combined to establish (D.1). For a sampled cartoon function $f \in \mathcal{C}^{N,K}_{\mathrm{CART}}$, i.e.,

$$\begin{aligned} f[n] &= c_1(n/N) + \mathbb{1}_{[a,b]}(n/N)c_2(n/N) \\ &=: f_1[n] + \mathbb{1}^N_{[a,b]}[n]f_2[n], \quad n \in I_N, \end{aligned}$$

we have

$$\begin{aligned} &\|f - F_\tau f\|_2 \le \|f_1 - F_\tau f_1\|_2 + \|\mathbb{1}^N_{[a,b]}(f_2 - F_\tau f_2)\|_2 \\ &+ \|(\mathbb{1}^N_{[a,b]} - F_\tau \mathbb{1}^N_{[a,b]})(F_\tau f_2)\|_2 \\ &\le \|f_1 - F_\tau f_1\|_2 + \|f_2 - F_\tau f_2\|_2 \\ &+ \|\mathbb{1}^N_{[a,b]} - F_\tau \mathbb{1}^N_{[a,b]}\|_2 \|F_\tau f_2\|_\infty, \end{aligned} \tag{D.4}$$

where in (D.4) we used

$$\begin{aligned} \big(F_\tau(\mathbb{1}^N_{[a,b]} f_2)\big)[n] &= (\mathbb{1}_{[a,b]}c_2)(n/N - \tau(n/N)) \\ &= \mathbb{1}_{[a,b]}(n/N - \tau(n/N))c_2((n/N - \tau(n/N))) \\ &= (F_\tau \mathbb{1}^N_{[a,b]})[n](F_\tau f_2)[n]. \end{aligned}$$

With the upper bounds (D.2) and (D.3), invoking properties of $\mathcal{C}^{N,K}_{\mathrm{CART}}$ (namely, (i) c_1, c_2 satisfy the Lipschitz property with Lipschitz constant $C = K$ and hence $f_1[n] = c_1(n/N), f_2[n] = c_2(n/N)$, $n \in I_N$, satisfy (D.2) with $C = K$, and (ii) $\|F_\tau f_2\|_\infty = \sup_{n \in I_N} |(F_\tau f_2)[n]| = \sup_{n \in I_N} |c_2(n/N - \tau(n/N))| \le \sup_{x \in \mathbb{R}} |c_2(x)| = \|c_2\|_\infty \le K$), this yields

$$\begin{aligned} \|f - F_\tau f\|_2 &\le 2KN^{1/2}\,\|\tau\|_\infty + 2KN^{1/2}\|\tau\|^{1/2}_\infty \\ &\le 4KN^{1/2}\|\tau\|^{1/2}_\infty, \end{aligned}$$

where in the last step we used $\|\tau\|_\infty \le \|\tau\|^{1/2}_\infty$, which is thanks to the assumption $\|\tau\|_\infty \le 1$. This completes the proof of (D.1). □

It remains to establish (D.2) and (D.3).

Lemma D.1. *Let $c : \mathbb{R} \to \mathbb{C}$ be Lipschitz-continuous with Lipschitz constant C. Let further $f[n] := c(n/N)$, $n \in I_N$. Then,*

$$\|f - F_\tau f\|_2 \le CN^{1/2}\|\tau\|_\infty.$$

Proof. Invoking the Lipschitz property of c according to

$$\begin{aligned} \|f - F_\tau f\|_2^2 &= \sum_{n \in I_N} |f[n] - (F_\tau f)[n]|^2 \\ &= \sum_{n \in I_N} |c(n/N) - c(n/N - \tau(n/N))|^2 \\ &\le C^2 \sum_{n \in I_N} |\tau(n/N)|^2 \le C^2 N \|\tau\|^2_\infty \end{aligned}$$

completes the proof. □

We continue with a deformation sensitivity result for sampled indicator functions $\mathbb{1}_{[a,b]}(x)$.

Lemma D.2. *Let $[a, b] \subseteq [0, 1]$ and set $\mathbb{1}^N_{[a,b]}[n] := \mathbb{1}_{[a,b]}(n/N)$, $n \in I_N$, with $a, b \notin \{0, \frac{1}{N}, \dots, \frac{N-1}{N}\}$. Then, we have*

$$\|\mathbb{1}^N_{[a,b]} - F_\tau \mathbb{1}^N_{[a,b]}\|_2 \le 2N^{1/2}\|\tau\|^{1/2}_\infty.$$

Proof. In order to upper-bound

$$\begin{aligned} \|\mathbb{1}^N_{[a,b]} - F_\tau \mathbb{1}^N_{[a,b]}\|^2_2 &= \sum_{n \in I_N} |\mathbb{1}^N_{[a,b]}[n] - (F_\tau \mathbb{1}^N_{[a,b]})[n]|^2 \\ &= \sum_{n \in I_N} |\mathbb{1}_{[a,b]}(n/N) - \mathbb{1}_{[a,b]}(n/N - \tau(n/N))|^2, \end{aligned}$$

we first note that the summand $h(n) := |\mathbb{1}_{[a,b]}(n/N) - \mathbb{1}_{[a,b]}(n/N - \tau(n/N))|^2$ satisfies $h(n) = 1$, for $n \in S$, where

$$S := \Big\{ n \in I_N \,\Big|\, \frac{n}{N} \in [a,b] \text{ and } \frac{n}{N} - \tau\Big(\frac{n}{N}\Big) \notin [a,b] \Big\} \cup \Big\{ n \in I_N \,\Big|\, \frac{n}{N} \notin [a,b] \text{ and } \frac{n}{N} - \tau\Big(\frac{n}{N}\Big) \in [a,b] \Big\},$$

and $h(n) = 0$, for $n \in I_N \backslash S$. Thanks to $a, b \notin \{0, \frac{1}{N}, \dots, \frac{N-1}{N}\}$, we have $S \subseteq \Sigma$, where

$$\Sigma := \Big\{ n \in \mathbb{Z} \,\Big|\, \Big|\frac{n}{N} - a\Big| < \|\tau\|_\infty \Big\} \cup \Big\{ n \in \mathbb{Z} \,\Big|\, \Big|\frac{n}{N} - b\Big| < \|\tau\|_\infty \Big\}.$$

The cardinality of the set Σ can be upper-bounded by $2\,\frac{2\|\tau\|_\infty}{1/N}$, which then yields

$$\|\mathbb{1}^N_{[a,b]} - F_\tau \mathbb{1}^N_{[a,b]}\|_2^2 = \sum_{n \in I_N} |h(n)|^2 = \sum_{n \in S} 1 \le \sum_{n \in \Sigma} 1 \le 4N\|\tau\|_\infty. \tag{D.5}$$

This completes the proof.

Remark D.2. *For general $a, b \in [0,1]$, i.e., when we drop the assumption $a, b \notin \{0, \frac{1}{N}, \dots, \frac{N-1}{N}\}$, it follows that $S \subseteq \Sigma'$, where*

$$\Sigma' := \Big\{ n \in \mathbb{Z} \,\Big|\, \Big|\frac{n}{N} - a\Big| \le \|\tau\|_\infty \Big\} \cup \Big\{ n \in \mathbb{Z} \,\Big|\, \Big|\frac{n}{N} - b\Big| \le \|\tau\|_\infty \Big\}.$$

Noting that the cardinality of Σ' can be upper-bounded by $2\big(\frac{2\|\tau\|_\infty}{1/N} + 1\big) = 4N\|\tau\|_\infty + 2$, this then yields (similarly to (D.5))

$$\|\mathbb{1}^N_{[a,b]} - F_\tau \mathbb{1}^N_{[a,b]}\|_2^2 \le \sum_{n \in \Sigma} 1 \le 4N\|\tau\|_\infty + 2,$$

which shows that the deformation error—for general $a, b \in [0,1]$—does not decay with $\|\tau\|_\infty^\alpha$ for some $\alpha > 0$ (see also the example in Remark D.1).

□

E. Appendix: Theorem 2

We start by establishing i). For ease of notation, again, we let $f_q := U[q]f$ and $h_q := U[q]h$, for $f, h \in H_{N_1}$, $q \in \Lambda_1^d$. We have

$$|||\Phi^d_\Omega(f) - \Phi^d_\Omega(h)|||^2 = \sum_{q \in \Lambda_1^d} ||(f_q - h_q) * \chi_d||_2^2 \tag{E.1}$$

$$\le \|\chi_d\|_1^2 \underbrace{\sum_{q \in \Lambda_1^d} ||(f_q - h_q)||_2^2}_{=: a_d}, \tag{E.2}$$

where (E.2) follows by Young's inequality (Folland, 2015).

Remark E.1. *We emphasize that (E.1) can also be upper-bounded by $B_{d+1} \sum_{q \in \Lambda_1^d} ||(f_q - h_q)||_2^2$, which follows from the fact that $\{g_{\lambda_{d+1}}\}_{\lambda_{d+1} \in \Lambda_{d+1}} \cup \{\chi_d\}$ are atoms of the convolutional set Ψ_{d+1} with Bessel bound B_{d+1}. Hence, one can substitute $\|\chi_d\|_1$ in (15) by B_{d+1}.*

The key step is then to show that a_d can be upper-bounded according to

$$a_k \le (B_k L_k^2 R_k^2) a_{k-1}, \qquad k = 1, \dots, d, \tag{E.3}$$

and to note that

$$a_d \le (B_d L_d^2 R_d^2) a_{d-1} \le \dots \le \Big(\prod_{k=1}^d B_k L_k^2 R_k^2 \Big) a_0 = \Big(\prod_{k=1}^d B_k L_k^2 R_k^2 \Big) \sum_{q \in \Lambda_1^0} \|f_q - h_q\|_2^2 = \Big(\prod_{k=1}^d B_k L_k^2 R_k^2 \Big) \|f - h\|_2^2,$$

which yields (16). We now establish (E.3). Every path

$$\tilde{q} \in \Lambda_1^k = \underbrace{\Lambda_1 \times \dots \times \Lambda_{k-1}}_{= \Lambda_1^{k-1}} \times \Lambda_k$$

of length k can be decomposed into a path $q \in \Lambda_1^{k-1}$ of length $k-1$ and an index $\lambda_k \in \Lambda_k$ according to $\tilde{q} = (q, \lambda_k)$. Thanks to (5) we have $U[\tilde{q}] = U[(q, \lambda_k)] = U_k[\lambda_k]U[q]$, which yields

$$\sum_{\tilde{q} \in \Lambda_1^k} \|f_{\tilde{q}} - h_{\tilde{q}}\|_2^2 = \sum_{q \in \Lambda_1^{k-1}} \sum_{\lambda_k \in \Lambda_k} \|U_k[\lambda_k] f_q - U_k[\lambda_k] h_q\|_2^2. \tag{E.4}$$

We next note that the term inside the sums on the RHS in (E.4) satisfies

$$\|U_k[\lambda_k] f_q - U_k[\lambda_k] h_q\|_2^2 = \|P_k\big(\rho_k(f_q * g_{\lambda_k})\big) - P_k\big(\rho_k(h_q * g_{\lambda_k})\big)\|_2^2 \le L_k^2 R_k^2 \|(f_q - h_q) * g_{\lambda_k}\|_2^2, \tag{E.5}$$

where we used the Lipschitz continuity of P_k and ρ_k with Lipschitz constants $R_k > 0$ and $L_k > 0$, respectively. As $\{g_{\lambda_k}\}_{\lambda_k \in \Lambda_k} \cup \{\chi_{k-1}\}$ are the atoms of the convolutional set Ψ_k, and $f_q, h_q \in H_{N_k}$ by (5), we have

$$\sum_{\lambda_k \in \Lambda_k} \|(f_q - h_q) * g_{\lambda_k}\|_2^2 \le B_k \|f_q - h_q\|_2^2,$$

which, when used in (E.5) together with (E.4), yields

$$\sum_{\tilde{q} \in \Lambda_1^k} \|f_{\tilde{q}} - h_{\tilde{q}}\|_2^2 \le B_k L_k^2 R_k^2 \sum_{q \in \Lambda_1^{k-1}} \|f_q - h_q\|_2^2,$$

and hence establishes (E.3), thereby completing the proof of i).

We now turn to ii). The proof of (17) follows—as in the proof of ii) in Theorem 1 in Appendix C—from (16) together with $\Phi^d_\Omega(h) = \{(U[q]h) * \chi_d\}_{q\in\Lambda_1^d} = 0$ for $h = 0$, see (C.13).

We continue with iii). The proof of the deformation sensitivity bound (18) is based on two key ingredients. The first one is the Lipschitz continuity result in (16). The second ingredient is, again, the deformation sensitivity bound (D.1) stated in Proposition D.1 in Appendix D. Combining (16) and (D.1)—as in the proof of iii) in Theorem 1 in Appendix C—then establishes (18) and completes the proof of iii).

We proceed to iv). For ease of notation, again, we let $f_q := U[q]f$, for $f \in H_{N_1}$, $q \in \Lambda_1^d$. Thanks to (5), we have $f_q \in H_{N_{d+1}}$, for $q \in \Lambda_1^d$. The key step in establishing (19) is to show that the operator U_k, $k \in \{1, 2, \dots, d\}$, defined in (4) satisfies the relation

$$(U_k[\lambda_k]T_m f) = T_{m/S_k}(U_k[\lambda_k]f), \tag{E.6}$$

for $f \in H_{N_k}$, $m \in \mathbb{Z}$ with $\frac{m}{S_k} \in \mathbb{Z}$, and $\lambda_k \in \Lambda_k$. With the definition of $U[q]$ in (5) this then yields

$$(U[q]T_m f) = T_{m/(S_1\cdots S_d)}(U[q]f), \tag{E.7}$$

for $f \in H_{N_1}$, $m \in \mathbb{Z}$ with $\frac{m}{S_1\dots S_d} \in \mathbb{Z}$, and $q \in \Lambda_1^d$. The identity (19) is then a direct consequence of (E.7) and the translation-covariance of the circular convolution operator (which holds thanks to $\frac{m}{S_1\dots S_d} \in \mathbb{Z}$):

$$\begin{aligned}\Phi^d_\Omega(T_m f) &= \big\{\big(U[q]T_m f\big) * \chi_d\big\}_{q\in\Lambda_1^d}\\ &= \big\{\big(T_{m/(S_1\cdots S_d)}U[q]f\big) * \chi_d\big\}_{q\in\Lambda_1^d}\\ &= \big\{T_{m/(S_1\cdots S_d)}\big((U[q]f) * \chi_d\big)\big\}_{q\in\Lambda_1^d}\\ &= T_{m/(S_1\cdots S_d)}\Phi^d_\Omega(f),\end{aligned}$$

for $f \in H_{N_1}$ and $m \in \mathbb{Z}$ with $\frac{m}{S_1\dots S_d} \in \mathbb{Z}$. It remains to establish (E.6):

$$\begin{aligned}(U_k[\lambda_k]T_m f) &= \Big(P_k\big(\rho_k((T_m f) * g_{\lambda_k})\big)\Big)\\ &= \Big(P_k\big(\rho_k(T_m(f * g_{\lambda_k}))\big)\Big) \qquad \text{(E.8)}\\ &= \Big(P_k\big(T_m(\rho_k(f * g_{\lambda_k}))\big)\Big), \qquad \text{(E.9)}\end{aligned}$$

where in (E.8) we used the translation covariance of the circular convolution operator (which holds thanks to $m \in \mathbb{Z}$), and in (E.9) we used the fact that point-wise non-linearities commute with the translation operator thanks to

$$\begin{aligned}(\rho_k T_m f)[n] &= \rho_k((T_m f)[n])\\ &= \rho_k(f[n-m]) = (T_m \rho_k f)[n],\end{aligned}$$

for $f \in H_{N_k}$, $n \in I_{N_k}$, and $m \in \mathbb{Z}$. Next, we note that the pooling operators P_k in Section 2.3.1 (namely, sub-sampling, average pooling, and max-pooling) can all be written as $(P_k f)[n] = (P'_k f)[S_k n]$, for some P'_k that commutes with the translation operator, namely, for (i) sub-sampling $(P'_k f)[n] = f[n]$, with $(P'_k T_m f)[n] = (T_m f)[n] = f[n-m] = (T_m P'_k f)[n]$, (ii) average pooling $(P'_k f)[n] = \sum_{l=n}^{n+S_k-1} \alpha_{l-n} f[l]$ with

$$\begin{aligned}(P'_k T_m f)[n] &= \sum_{l=n}^{n+S_k-1} \alpha_{l-n} f[l-m]\\ &= \sum_{l'=(n-m)}^{(n-m)+S_k-1} \alpha_{l-(n-m)} f[l']\\ &= (T_m P'_k f)[n],\end{aligned}$$

and for (iii) max-pooling $(P'_k f)[n] = \max_{l\in\{n,\dots,n+S_k-1\}} |f[l]|$ with

$$\begin{aligned}(P'_k T_m f)[n] &= \max_{l\in\{n,\dots,n+S_k-1\}} |f[l-m]|\\ &= \max_{(l-m)\in\{n-m,\dots,(n-m)+S_k-1\}} |f[l-m]|\\ &= \max_{l'\in\{(n-m),\dots,(n-m)+S_k-1\}} |f[l']|\\ &= (T_m P'_k f)[n],\end{aligned}$$

in all three cases for $f \in H_{N_k}$, $n \in I_{N_k}$, and $m \in \mathbb{Z}$. This then yields

$$\begin{aligned}(P_k T_m f)[n] &= (P'_k T_m f)[S_k n] = (T_m P'_k f)[S_k n]\\ &= P'_k(f)[S_k n - m]\\ &= P'_k(f)[S_k(n - S_k^{-1}m)]\\ &= P_k(f)[n - S_k^{-1}m]\\ &= (T_{m/S_k} P_k f)[n], \end{aligned}\tag{E.10}$$

for $f \in H_{N_k}$ and $n \in I_{N_{k+1}}$. Here, we used $m/S_k \in \mathbb{Z}$, which is by assumption. Substituting (E.10) into (E.9) finally yields

$$(U_k[\lambda_k]T_m f) = T_{m/S_k} U_k[\lambda_k]f,$$

for $f \in H_{N_k}$, $m \in \mathbb{Z}$ with $\frac{m}{S_k} \in \mathbb{Z}$, and $\lambda_k \in \Lambda_k$. This completes the proof of (E.6) and hence establishes (19).

References

Folland, G. B. *A course in abstract harmonic analysis*, volume 29. CRC Press, 2015.

Golub, G. H. and Van Loan, C. F. *Matrix computations*. Johns Hopkins University Press, 2013.

Wiatowski, T. and Bölcskei, H. A mathematical theory of deep convolutional neural networks for feature extraction. *arXiv:1512.06293*, 2015.