

Deep learning in turbulent convection networks

Enrico Fonda^{a,1}, Ambrish Pandey^{b,1}, Jörg Schumacher^{b,c}, and Katepalli R. Sreenivasan^{a,c,d,e,2}

^aDepartment of Physics, New York University, New York, NY 10012; ^bDepartment of Mechanical Engineering, Technische Universität Ilmenau, D-98684 Ilmenau, Germany; ^cTandon School of Engineering, New York University, New York, NY 11201; ^dCourant Institute of Mathematical Sciences, New York University, New York, NY 10012; and ^eSchool of Mathematics, Institute for Advanced Study, Princeton, NJ 08540

Contributed by Katepalli R. Sreenivasan, March 11, 2019 (sent for review January 9, 2019; reviewed by Michael Chertkov and Karthik Duraisamy)

We explore heat transport properties of turbulent Rayleigh–Bénard convection in horizontally extended systems by using deep-learning algorithms that greatly reduce the number of degrees of freedom. Particular attention is paid to the slowly evolving turbulent superstructures—so called because they are larger in extent than the height of the convection layer—which appear as temporal patterns of ridges of hot upwelling and cold downwelling fluid, including defects where the ridges merge or end. The machine-learning algorithm trains a deep convolutional neural network (CNN) with U-shaped architecture, consisting of a contraction and a subsequent expansion branch, to reduce the complex 3D turbulent superstructure to a temporal planar network in the midplane of the layer. This results in a data compression by more than five orders of magnitude at the highest Rayleigh number, and its application yields a discrete transport network with dynamically varying defect points, including points of locally enhanced heat flux or “hot spots.” One conclusion is that the fraction of heat transport by the superstructure decreases as the Rayleigh number increases (although they might remain individually strong), correspondingly implying the increased importance of small-scale background turbulence.

turbulent convection | machine learning | temporal networks

Novel machine-learning algorithms (1–4), particularly in the form of multilayered (or deep) neural networks, have lately found various applications in the big data domain, but are yet to find an established place in the study of turbulent flows despite a few promising beginnings (5–7). In this paper, we apply machine-learning tools to detect flow patterns in spatially extended turbulent convection and explore their fluid-dynamic significance. The data needed for training the network and detecting flow patterns both come from the full-scale simulation of the equations of motion (although from independent segments). Indeed, the continued increase of supercomputing power in recent years has enabled the direct numerical simulations of highly turbulent flows (8–10), resolving the entire array of scales from the dissipation scales to superstructures in the form of ordered patterns larger than the fluid flow layer thickness itself (10–15).

The output of massive simulations is turbulent fields that require for proper assimilation a change of paradigms in data processing, structure recognition, and subgrid modeling. Supervised machine learning makes use of the fact that it is often easier to train a system, such as a deep convolutional neural network in the present case, with a number of labeled examples of an intended input–output than to develop a specific computer program to provide the correct answer for all possible input data.

In this work, we apply deep-learning methods to reduce the 3D and fully resolved superstructures in the temperature field of a Rayleigh–Bénard flow (10, 15) to a 2D planar temporal network (16, 17) of maximal and minimal temperature ridges with point defects. Prominent and persistent large-scale patterns are known to exist in many convection flows in nature, e.g., cloud streets in the atmosphere of the Earth (18) or as granulation and supergranulation on the solar surface (19, 20).

Our goal is to quantify the fraction and time variability of the turbulent heat transfer associated with superstructures by means of a vastly smaller dataset via a machine-learning algorithm that trains a specific deep convolutional neural network (CNN). This U-shaped network (U-net for short) combines a contraction path of a standard CNN (21) with a subsequent expansion path of concatenations and up-convolutions that finally creates a detailed segmentation map (22). The U-net has been applied successfully to image segmentation of touching objects, e.g., of neuronal cell structures in electron microscopy data (23), and advanced very recently fluorescence microscopy to new limits (24). The specific architecture of the U-net is essential for the present application as it requires smaller sets of manually annotated data for the training, in comparison with standard cases for overlapping objects. The slow evolution of the superstructures in the large-aspect-ratio domain would otherwise require extremely long simulations of the fully 3D turbulent flow over a few tens of thousands of convective time units to obtain an appropriate amount of independent training data, as discussed in ref. 10.

Turbulent superstructures in convection display a large-scale order that is obtained by a time-windowed averaging of the velocity and temperature fields in the Eulerian frame (10, 25, 26) while instantaneous snapshots of these quantities are fully turbulent as described in *SI Appendix*. These superstructures remind us of convection rolls and temperature ridges in the weakly nonlinear regime above the onset of convection (27, 28). Similar to weakly nonlinear structures, superstructures have defects, i.e., local imperfections in their arrangement, as is visible in Fig. 1 for different physical quantities for three Rayleigh numbers. Local maxima and minima of time-averaged temperature Θ , which is later defined precisely, are arranged in the form of ridges of

Significance

Turbulent convection in horizontally extended systems comprises vortices and plumes on many time and length scales. These structures interact nonlinearly to self-organize into slowly evolving turbulent superstructures, which are horizontally more extended than in height. We use a U-shaped deep-learning algorithm to generate a time-varying planar network, resulting in a drastic reduction of degrees of freedom, and use it to detect the 3D superstructures and estimate their effectiveness in transporting heat. We thus demonstrate the likely utility of deep learning for parameterizing convection in global models of atmospheric and stellar convection whenever mesoscale structures are conspicuous.

Author contributions: J.S. and K.R.S. designed research; E.F., A.P., and J.S. performed research; E.F. and A.P. analyzed data; and E.F., A.P., J.S., and K.R.S. wrote the paper.

Reviewers: M.C., University of Arizona; and K.D., University of Michigan.

The authors declare no conflict of interest.

Published under the [PNAS license](#).

¹E.F. and A.P. contributed equally to this work.

²To whom correspondence should be addressed. Email: katepalli.sreenivasan@nyu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1900358116/-DCSupplemental.

Published online April 15, 2019.

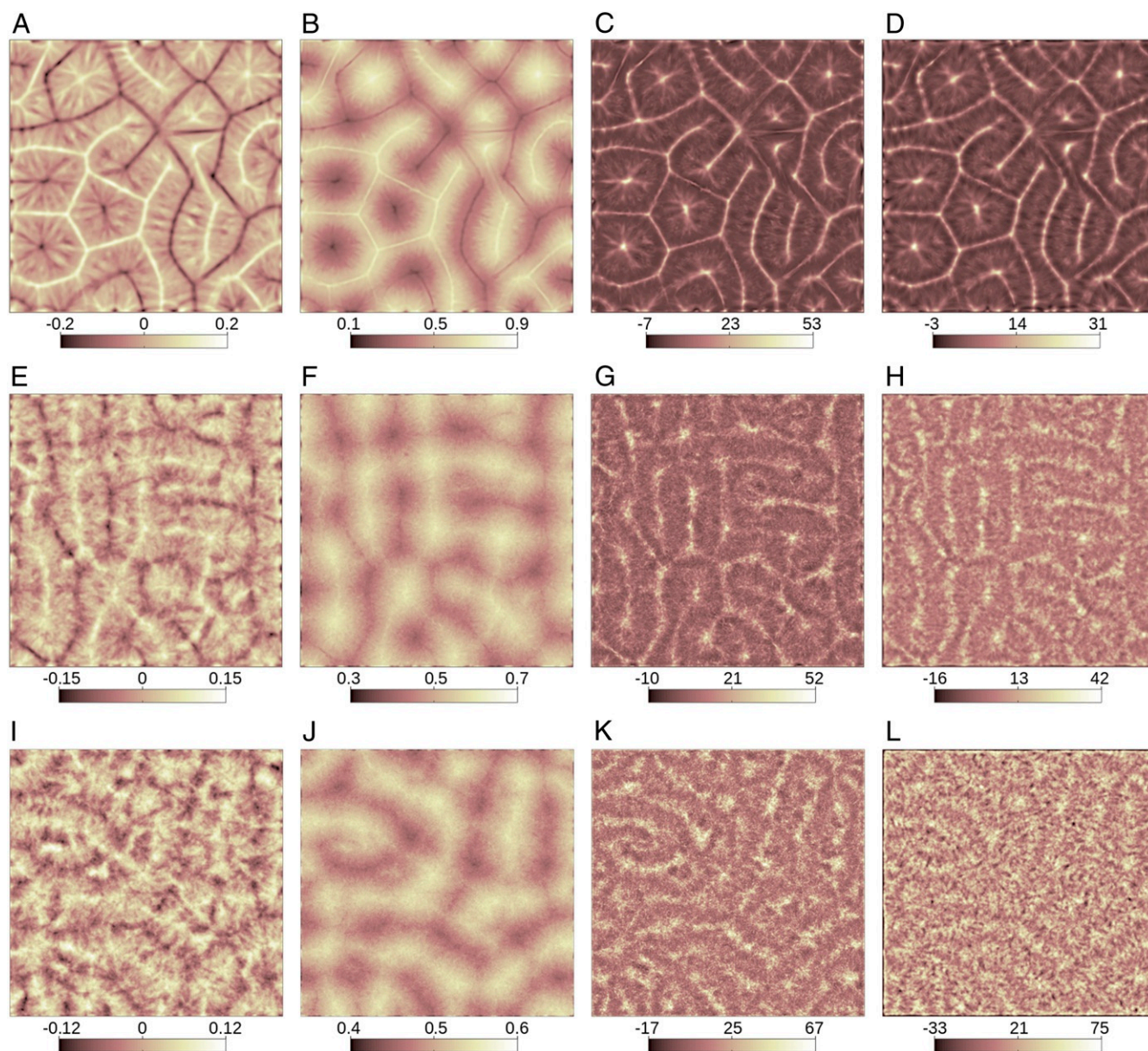


Fig. 1. Turbulent superstructures for three Rayleigh–Bénard flows at Prandtl number $Pr = 7$. Data are obtained by windowed averaging over a time interval τ , as discussed in the main text. Data in *A–D* are for a Rayleigh number $Ra = 10^5$, data in *E–H* are for $Ra = 10^6$, and data in *I–L* are for $Ra = 10^7$. (*A*, *E*, and *I*) Vertical velocity component $U(x, y)$ in the midplane at $z = 1/2$. (*B*, *F*, and *J*) Temperature $\Theta(x, y)$ in the midplane. (*C*, *G*, and *K*) Local Nusselt number $Nu_l(x, y)$ in the midplane. All three quantities are defined by Eq. 1. *D*, *H*, and *L* show, in addition, the vertically averaged heat flux. Color bars below the panels indicate the ranges for the corresponding data. The sides are 25 times the height.

hot upwelling and cold downwelling fluid in the midplane at $z = H/2$, where H is the layer height (Fig. 1 *B*, *F*, and *J*). Ridges end abruptly in wedge defect points; two ridges merge into one at trisector defect points. These ridges coincide with the local extrema of the time-averaged vertical velocity U in Fig. 1 *A*, *E*, and *I* or the time-averaged local Nusselt number Nu_l in Fig. 1 *C*, *G*, and *K*, even after an additional averaging over the whole layer height H shown in Fig. 1 *D*, *H*, and *L*. Since all these fields provide the same visual information, we conduct the image segmentation for the simplest measurable quantity, the temperature. We verified that vertically averaged temperature snapshots provide the same input as midplane slices.

Fig. 1 *B*, *F*, and *J* shows further that the temperature ridges form a network that connects the defect points. In other words,

the key elements of the turbulent superstructure can be assembled into a network $G(n, e, w)$ consisting of a finite set of nodes n (= defects) and edges e (= ridges) with a certain weight w (16). With increasing Rayleigh number, the turbulent data, even when averaged over a time window τ , remain noisy as seen in Fig. 1 *E–L*; this means that standard techniques for the detection of defects in weakly nonlinear Rayleigh–Bénard convection, which are based on phase field analysis (29) or local wave vectors (30), generate many false positives. The same holds for standard edge detection algorithms as demonstrated in *SI Appendix*. On the other hand, the present U-net provides reliable data reduction to the ridge network: For the highest Rayleigh number studied here, we obtain a data compression by a factor of 0.2 million, from 13.5 GB for a superstructure snapshot to 65.5 kB for the network. Our work thus combines deep learning

Table 1. Parameters of the different spectral element simulations conducted for the superstructure analysis

Quantity	Run 1T	Run 1	Run 2T	Run 2	Run 3
Ra	10^5	10^5	10^6	10^6	10^7
Nu	4.13 ± 0.01	4.16 ± 0.01	8.30 ± 0.02	8.30 ± 0.02	16.2 ± 0.04
u_{rms}/U_f	0.089 ± 0.001	0.090 ± 0.001	0.102 ± 0.001	0.102 ± 0.001	0.109 ± 0.001
Re	10.7 ± 0.1	10.7 ± 0.1	38.7 ± 0.2	38.7 ± 0.2	130 ± 0.5
λ_U/H	5.6	5.4	5.7	5.7	4.3
λ_Θ/H	6.3	6.3	6.3	6.3	6.2
τ_{total}/T_f	2,670	2,315	940	2,093	471
τ/T_f	207	198	179	179	138
t_0/T_f	10.0	2.50	5.44	5.44	2.98
N_s	268	927	174	386	159
N_τ	12	848	12	527	113

For all examples $Pr = 7$ and $\Gamma = 25$. Runs 1T and 2T are the training data runs, and runs 1, 2, and 3 provide the analysis data (or test data). We list Rayleigh number Ra, Nusselt number Nu, root-mean-square velocity u_{rms} , and the Reynolds number Re. Furthermore, we display the characteristic scales λ_U and λ_Θ of vertical velocity and temperature, the total integration time τ_{total} , and the characteristic time τ of the superstructures, as determined in ref. 10. Finally, the number of original simulation snapshots, N_s , which are equidistantly distributed over the total time interval, and the number of τ -averaged temperature slices, N_τ , that serve as training or test data, are given. The output interval is given by the time $t_0 = \tau_{total}/(N_s - 1)$. Note that $N_\tau > N_s$ for run 2 since additional data from the training-set run 2T have been taken to get a longer time series.

with network theory to approximate turbulent heat transport by superstructures to a dynamically evolving, and much sparser, planar network in the symmetry plane of the convection layer. This approach opens simple ways for modeling turbulent transport on mesoscales of natural convection systems.

Turbulent Convection Flow

We solve in three space dimensions and time the coupled equations of motion for velocity field u_i and temperature T in the Boussinesq approximation of thermal convection (31) by direct numerical simulations using a spectral element method. The geometry is a closed cuboid cell with a square cross-section in horizontal directions, $V = L^2H$, and an aspect ratio $\Gamma = L/H = 25$, with L being the horizontal box length (10, 32, 33). The Rayleigh number Ra quantifies the thermal driving of the flow and the Prandtl number Pr the ratio of kinematic viscosity to temperature diffusivity. We use three datasets, all for $Pr = 7$, for Rayleigh numbers of $Ra = 10^5$, 10^6 , and 10^7 . Further details on the simulation method, the specific runs, and the boundary conditions are found in *SI Appendix* and in Table 1. The turbulent superstructure fields are obtained by a time-windowed averaging over time τ in the Eulerian frame of reference and are given by

$$\Phi_\tau(x, y, t_0) = \frac{1}{\tau} \int_{t_0 - \tau/2}^{t_0 + \tau/2} \phi(x, y, z_0, t') dt' \quad [1]$$

with $\Phi_\tau = \{U, \Theta, Nu_l\}$ and $\phi = \{u_z, T, \sqrt{RaPr} u_z \theta\}$, respectively. In this analysis, we decompose the temperature field into $\theta(\mathbf{x}, t) = T(\mathbf{x}, t) - T_{lin}(z)$ with $T_{lin}(z)$ as the linear temperature profile across the quiescent layer. The averaging time scale τ and the associated characteristic length scale λ separate the slow, large-scale from the fast, small-scale motion. These scales are summarized in Table 1, and more details on their calculation are described in ref. 10 and in *SI Appendix*. All quantities are expressed in units of the cell height H and the free-fall velocity $U_f = \sqrt{g\alpha\Delta TH}$, where g is the acceleration due to gravity, α the isobaric expansion coefficient, and ΔT the imposed temperature difference between the bottom and top plates. Consequently, times in Table 1 are given in units of $T_f = H/U_f$.

The flow responds to the input parameters Ra, Pr, and Γ with a turbulent heat and momentum transport quantified by

the (global) Nusselt number $Nu = 1 + \sqrt{RaPr} \langle u_z T \rangle_{V,t}$ and the Reynolds number $Re = \sqrt{Ra/Pr} u_{rms}$, respectively. The root-mean-square velocity is determined by combining volume and time averaging of all three velocity components.

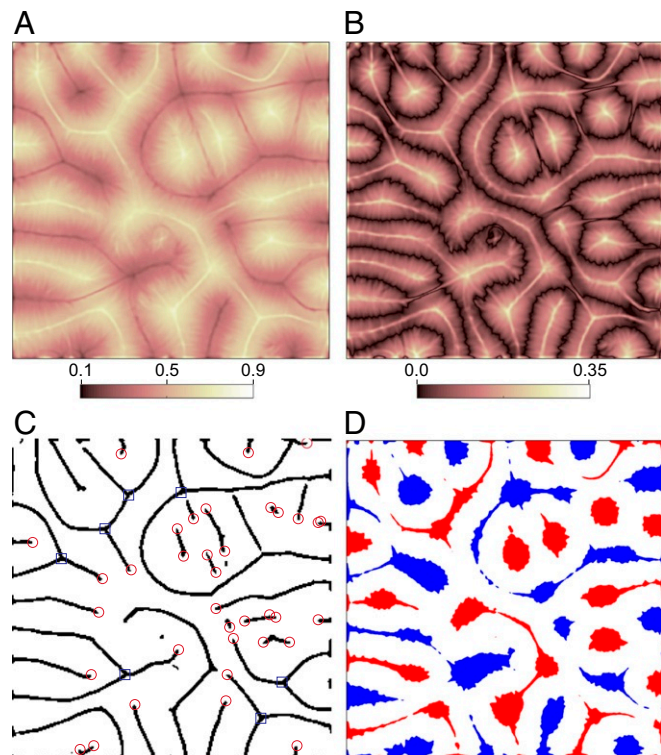


Fig. 2. Input and output of the U-net. (A) Original 2D input data slice of time-averaged temperature field Θ calculated by Eq. 1. (B) Symmetrized temperature field $\hat{\Theta}$ as calculated by Eq. 2. (C) Output of the U-net is a ridge pattern. The defect points that are subsequently detected are also indicated (wedge points, red circles; trisectors, blue squares). (D) Comparison with a simple thresholding set by blue pixels for $T < 0.385$ and red pixels for $T > 0.615$. As is visible, the ridges have a strongly varying thickness and are partly interrupted when a simple thresholding is imposed.

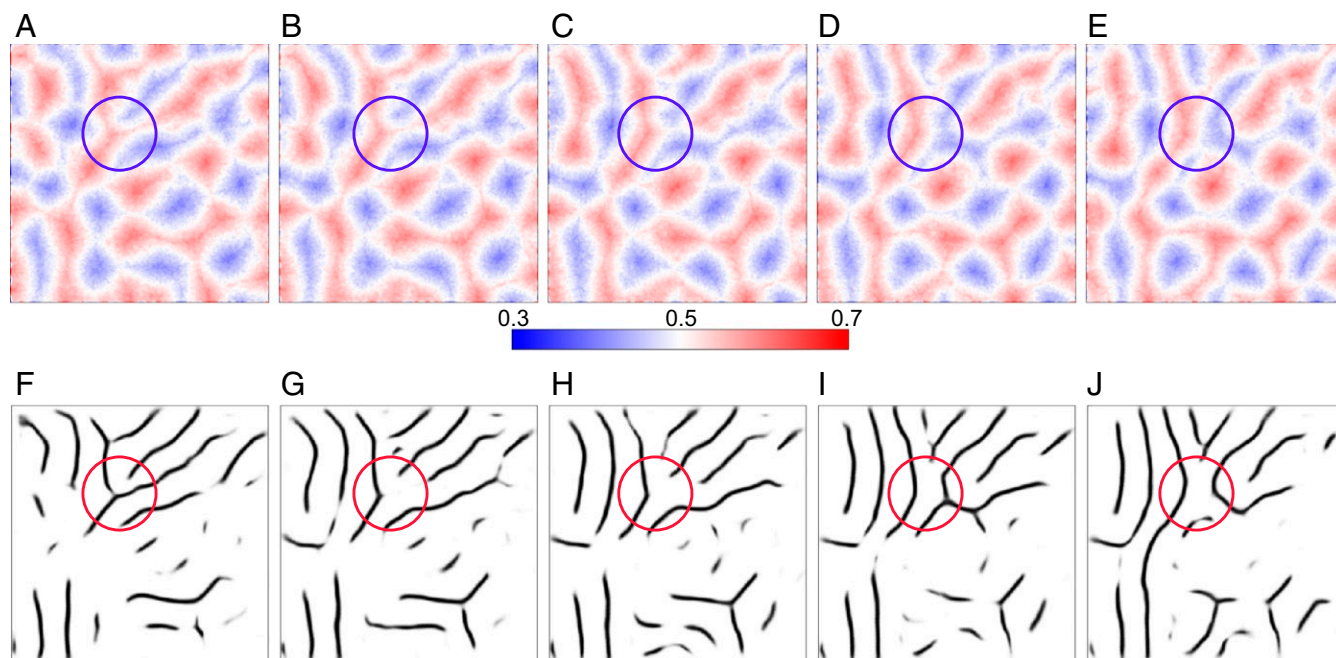


Fig. 3. Slow evolution of the ridge network with annihilation and generation of pattern defect points. *A–E* show the temperature field in the midplane. *F–J* show the corresponding output of the U-net. *A* and *F* correspond to $t/T_f = 918$ ($=5.13\tau$), *B* and *G* to $t/T_f = 972$ ($=5.43\tau$), *C* and *H* to $t/T_f = 1,026$ ($=5.73\tau$), *D* and *I* to $t/T_f = 1,081$ ($=6.04\tau$), and *E* and *J* to $t/T_f = 1,135$ ($=6.34\tau$). The color bar is the same for *A–E*. Data are from run 2 for $Ra = 10^6$.

Data Reduction by CNN

The present U-net (22) applies local convolution operations between (neighboring) neurons in 26 hidden layers, thus consisting of 28 layers when the input and output layers are added. Input of the U-net is τ -averaged 2D temperature slices which are taken in the midplane and symmetrized:

$$\hat{\Theta}(x, y, t_0) = \left| \Theta(x, y, t_0) - \frac{1}{2} \right|. \quad [2]$$

In this way, hot and cold ridges of the temperature field become the same type of features to be extracted. The present CNN consists of a standard contraction path as in all standard convolutional networks. In each hidden layer, 3×3 convolutions of the neighboring pixels are followed by a nonlinear activation with an exponential linear unit (ELU) describing how to fire each neuron in the U-net (34). In a subsequent 2×2 max-pooling operation the number of pixels is reduced by a factor of 2 in each direction. Such an operation reduces the computational cost and makes the whole CNN more resilient to noise. The sequence of operations reduces the resolution of the input image from 256^2 to 16^2 points via successive layers, and via convolutions there is an increase of the number of feature maps (images) from 1 to 256. The expansion path of the neural network, which is the specific new feature of the U-net (22), passes the levels from lower to higher resolution by up-convolutions and concatenations with data from the contraction branch. A detailed description of the data processing across the U-net is found in *SI Appendix*.

Our goal is to segment the simulation data images into two sets: pixels that belong to the temperature ridge skeleton and the rest. Essential for the success of this goal is the learning algorithm. The weights of the neurons of the U-net are initially small random numbers. They are adjusted iteratively by means of training data—a separate subset of our simulation data which we denote with runs 1T or 2T, each with 12 manually annotated images—until a loss (or cost) functional is minimized. This is done within cycles or epochs amounting to about 250 in the

present cases. Data are propagated forward through the hidden layers of the deep neural network from the input to the output layer. The error is then propagated backward to the input layer to readjust the weights and start a forward propagation of the data anew. As seen from Table 1, the number of training data slices, N_t , is very small for the runs 1T and 2T—one big advantage of the present U-net compared with most other CNNs. We augment the training dataset by means of rotations and reflections by a factor of 8. Further details on the training algorithm of the U-net are found in *SI Appendix*. Also note that data of run 3 are analyzed with the same network as those of run 2. Output of the U-net is a partly connected 2D skeleton of ridges, which is then used to detect trisectors and wedge point positions. This skeleton can be thought as a network or a graph $G(n, e, w)$ with nodes (n) and edges (e) that carry a weight (w). This network decays into subnetworks and the distance between the ridges corresponds with the characteristic spatial scales of the original superstructure, λ .

Generation and Annihilation of Pattern Defects

Fig. 2 illustrates the essentials of data processing. The original temperature data slice $\Theta(x, y, 1/2, t_0)$ (Fig. 2*A*) is symmetrized to $\hat{\Theta}(x, y, 1/2, t_0)$, the U-net input. The output is shown in Fig. 2*C*—the ridge skeleton for which we detect the wedge and trisector points subsequently. The defect points are indicated by the symbols in Fig. 2*C*. Also, not all defect points are found by the U-net, as shown in Fig. 2; in other cases false positives are

Table 2. Defect point statistics of the simulation runs 1, 2, and 3 of Table 1

Ra	$\bar{T} \pm \sigma_T$	$\max(T)/\min(T)$	$\bar{W} \pm \sigma_W$	$\max(W)/\min(W)$
10^5	10 ± 2	18/5	41 ± 5	58/28
10^6	1 ± 1	4/0	32 ± 5	47/12
10^7	1 ± 1	2/0	32 ± 4	43/24

The mean numbers of trisectors (T) and wedge points (W) are given with error bars, as are the maximum and minimum numbers.

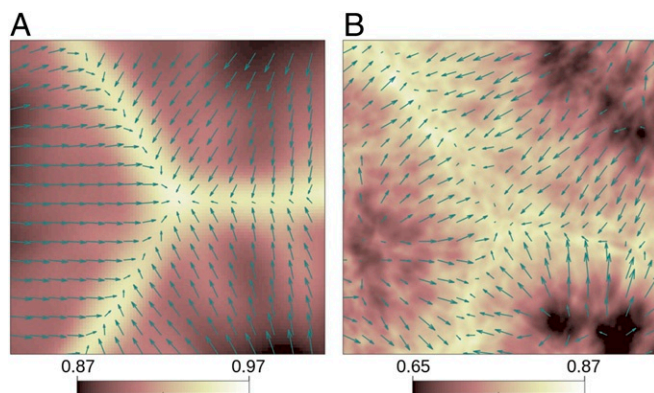


Fig. 4. Magnification of a trisector point region of a superstructure snapshot near the bottom plate. (A) Temperature (background) and horizontal velocity field (blue arrows) for $Ra = 10^5$ (run 1) at $z = H/30$ covering a square section of $1.76H$. (B) Same data for $Ra = 10^6$ (run 2) at $z = H/30$ with a side length of $4.89H$. All fields are averaged over the time interval τ .

detected. To measure the success rate of the ridge detection of the U-net we use the same standard metric used in segmentation problems, the intersection over union (IOU). The IOU is the ratio of true positives to the sum of true positives, false positives, and false negatives. This corresponds to the intersection between the ground truth (the images traced manually) and the prediction, divided by the union of the two. For the case of $Ra = 10^5$ the IOU is 0.57, for $Ra = 10^6$ it is 0.36, and for $Ra = 10^7$ it is 0.32. An estimate of this additional postprocessing error is given in *SI Appendix*. Even though the IOU is substantially smaller than unity, the detection of the ridge skeleton by the U-net is more reliable in terms of picking up the ridge patterns compared with a simple thresholding that is displayed in Fig. 2D. In *SI Appendix*, we show that the U-net also outperforms a classical edge detection algorithm and gives reliable output for the highest Ra where the superstructures are much noisier than the data in Fig. 2.

Fig. 3 illustrates the time evolution of the defect point network. Fig. 3A–E shows the temperature field Θ which is plotted together with the corresponding network output in Fig. 3F–J. The sequence spans a time interval of more than 200 free-fall time units, thus underlining the rather slow evolution of superstructures. The number of defect points is changing due to continued ongoing change of the ridge network topology by the annihilation of the old defect points and the formation of new ones. The number of defect points fluctuates in time. In Table 2, we summarize the findings for the three direct numerical simulation runs. It is seen that the mean number of trisectors decreases significantly from $Ra = 10^5$ to 10^6 whereas the mean number of wedge points remains about the same. This indicates that the polygon-type cells become fewer as Ra grows; this suggests a change of topology of the superstructures from closed cells to open rolls.

Heat Transfer Network by Superstructures

Heat is carried across the bulk region of the convection layer almost solely by correlations of vertical velocity component and temperature fluctuations. In other words, the local Nusselt number $Nu_l(x, y, t)$, defined in Eq. 1, is the central quantity for us (Fig. 1C, G, and K). The turbulent heat transport due to the network $G(n, e, w)$, extracted by the CNN from the turbulent superstructure (TSS), is given by a dimensionless Nusselt number,

$$Nu^{TSS}(t) = \sqrt{RaPr} [w_e \langle u_z \theta \rangle_e]. \quad [3]$$

The area fraction of the ridges (or edges e) is captured in the corresponding weight factor w_e . The average $\langle u_z \theta \rangle_e$ in Eq. 3 stands

for the local heat flux across an edge segment of unit length which is determined for a few representative snapshots from the original simulation data.

It is apparent in Fig. 1 (particularly Fig. 1C and D or G and H) that the local Nusselt number Nu_l is larger at the trisector points in comparison with that at the ridges (or edges of $G(n, e, w)$) and at the wedge points. These trisectors thus form local hot spots of the turbulent heat transport. The reason for such locally enhanced heat flux is illustrated in Fig. 4, where the vicinity of superstructure trisectors is displayed near the bottom wall for two Rayleigh numbers. The coincidence of three line-like plume ridges at the defect is connected with pronounced horizontally convergent flow parallel to the wall—forming a stable node structure that is a sink because $\partial u_x / \partial x + \partial u_y / \partial y < 0$. The convergent fluid flow is redirected vertically due to incompressibility, thus carrying hot fluid rapidly upward, with $\partial u_z / \partial z > 0$, and consequently enhancing the local heat and momentum transport.

For the lowest Rayleigh number, isolated hot spots in the form of very short ridges are detected in addition. They are traced back to strong plume impact regions at the opposite plates. However, their contribution to the global transport remains negligible since they are strongly localized. To summarize, the whole time dependence of the convective heat transport across the ridge network follows from the temporal change of its conformation only—i.e., from the area fraction in the plane that is covered by the network and the number of trisectors, respectively. The transport due to the 3D turbulent superstructure is reduced to that of a dynamic planar network.

Fig. 5 summarizes heat transport properties of the ridge network. Fig. 5A compares the (standard) Nusselt number obtained from averaged heat flux across the whole midplane,

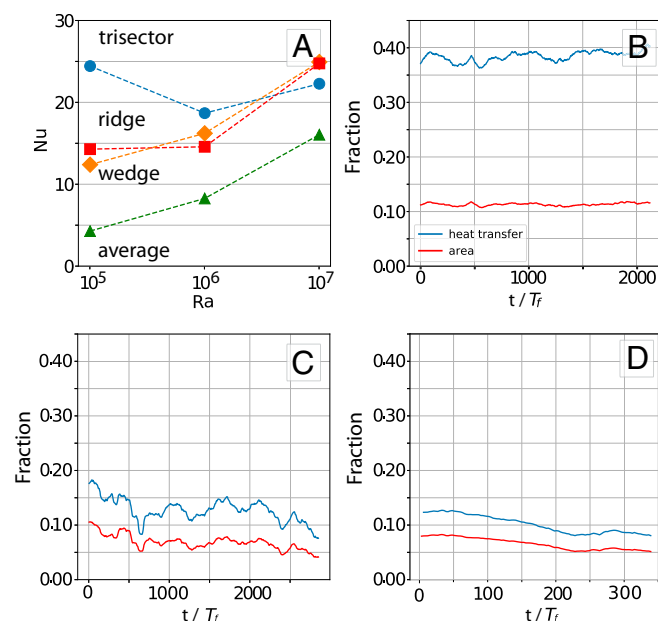


Fig. 5. Turbulent heat transport across the ridge network. (A) Comparison of (standard) mean convective heat flux across the whole layer (green) with that due to all trisector points (blue), all of the wedge points (orange), and the remaining ridge network (red), for varying Rayleigh numbers. (B–D) Amount of heat flux across the network including hot spots vs. time and area fraction of the network vs. time. Data are for run 1 at $Ra = 10^5$ in B, run 2 at $Ra = 10^6$ in C, and run 3 at $Ra = 10^7$ in D. The heat flux by the hot spots decreases with Ra but the area occupied by them also diminishes, suggesting that the hot spots individually remain efficient. The relatively smooth nature of curves in D, in contrast to B and C, is the effect of decreased data samples.

